Device-independent certification of one-shot distillable entanglement

Originally published in:
New Journal of Physics 21, https://doi.org/10.1088/1367-2630/aafef6

Funding acknowledgement:
- Information-theoretic methods for physics ()

PAPER • OPEN ACCESS

Device-independent certification of one-shot distillable entanglement

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
28 May 2018

REVISED
14 December 2018

ACCEPTED FOR PUBLICATION
16 January 2019

PUBLISHED
15 March 2019

PAPER

Device-independent certification of one-shot distillable entanglement

Rotem Arnon-Friedman[1,3] **and Jean-Daniel Bancal**[2]

[1] Institute for Theoretical Physics, ETH-Zürich, CH-8093, Zürich, Switzerland
[2] Department of Physics, University of Basel, CH-4056, Basel, Switzerland
[3] Author to whom any correspondence should be addressed.

E-mail: rotemaf@berkeley.edu

Keywords: entanglement, device-independent, quantum information, non-locality

Abstract

Entanglement sources that produce many entangled states act as a main component in applications exploiting quantum physics such as quantum communication and cryptography. Realistic sources are inherently noisy, cannot run for an infinitely long time, and do not necessarily behave in an independent and identically distributed manner. An important question then arises—how can one test, or certify, that a realistic source produces high amounts of entanglement? Crucially, a meaningful and operational solution should allow us to certify the entanglement which is available for further applications after performing the test itself (in contrast to assuming the availability of an additional source which can produce more entangled states, identical to those which were tested). To answer the above question and lower bound the amount of entanglement produced by an uncharacterised source, we present a protocol that can be run by interacting classically with uncharacterised (but not entangled to one another) measurement devices used to measure the states produced by the source. A successful run of the protocol implies that the remaining quantum state has high amounts of one-shot distillable entanglement. That is, one can distill many maximally entangled states out of the single remaining state. Importantly, our protocol can tolerate noise and, thus, certify entanglement produced by realistic sources. With the above properties, the protocol acts as the first 'operational device-independent entanglement certification protocol' and allows one to test and benchmark uncharacterised entanglement sources which may be otherwise incomparable.

1. Introduction

Entanglement is one of the most fundamental concepts of quantum physics, distinguishing it from classical physics [HHHH09]. Furthermore, it plays a crucial role in the advantages gained by considering applications of quantum physics such as quantum computation [Wil10], communication [DW02], and cryptography [BS16].

For most applications utilising entanglement, a single entangled pair of particles, e.g. a maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, is not sufficient. Instead, one must use many copies of entangled states or, to put differently, a highly entangled state, such as the maximally entangled state $|\Phi^L\rangle = \frac{1}{\sqrt{L}}\sum_{i=1}^{L}|i\rangle|i\rangle$ of rank L. Sources that produce high amounts of entanglement are, thus, a prerequisite for gaining from the computational and cryptographic advantages that quantum physics and quantum information unveil.

Three interesting questions then arise. Firstly, how should one *quantify* the amount of entanglement produced by a source? Secondly, how can one *compare*, or *benchmark*, different types of entanglement sources? Thirdly, how can one test, or *certify*, that high amounts of entanglement are indeed being produced by the source?

Importantly, we would like to answer these questions in an operational way. That is: (a) the suggested answer should be relevant for realistic sources and experimental settings. (b) When considering possible certification procedures, the statement should apply to the entanglement which is still present after the test rather than the entanglement which was already used and, hence, destroyed by the performed certification procedure itself.

The current work is concerned with certifying entanglement in such an operational way. Realistic sources are inherently noisy: they produce at best noisy entangled states. Moreover, a source creating many entangled pairs might produce systems that are correlated with one another due to, e.g. drifting of the source with time or the presence of a memory inside the source. In other words, the overall state produced by the source (after using it many times) is not an independent and identically distributed (IID) state. We call such sources *noisy non-IID sources*.

Clearly, in order to make any estimation of the quality of the source one must collect some data regarding the produced entanglement by performing certain measurements on the quantum states produced by the source. Similarly to the considerations regarding the source, the measurement devices might also behave in a noisy and non-IID manner.

1.1. Device-independent entanglement certification (DIEC)

We fix as a target to answer our questions while accommodating the noisy non-IID nature of the problem in the so called *device-independent* (DI) framework [Sca12]. In the DI approach one treats the quantum apparatuses as black boxes with which we can only interact classically. That is, we assume no prior knowledge regarding the internal behaviour of the source and the measurement devices. Assumptions are made, however, on the relation that certain devices can have with each other, and on the communication allowed between them. Concretely, we only interact with those devices by 'pushing buttons' and collecting the classical data output by them. For example, we may push a button on the source apparatus to produce a state and then another button on the measurement device to ask it to perform a measurement with 'input' 0 or 1. The state produced by the source is uncharacterised and we do not know which measurements are actually being performed when we use the inputs 0 and 1.

As widely known, the only way to demonstrate that the actions of physical devices cannot be explained by classical physics in a DI manner is to perform some Bell tests using the devices and observe a violation of the considered Bell inequality [Bel64, BCP+14]. A Bell violation acts as a 'witness' attesting to the quantum (in fact, non-local) nature of the devices used to violate the inequality. This can then be used to derive conclusions regarding, e.g. the structure of the underlying quantum states [SW87, PR92, CGS16] or the randomness of the measurements' outcomes [Col09, VV12].

We remark that by using the DI method we do not only treat imperfections in the quantum apparatuses which are known or can be characterised in advance. The DI approach allows one to derive conclusions without making assumptions regarding the types of imperfections. Even more drastically, the devices can be assumed to be 'malicious'[4]; as long as a violation of a Bell inequality is observed, the observer, or verifier, can be sure of the quantum nature of the systems without placing significant trust in the manufacture of the devices.

Most previous works, both theoretical and experimental, that can be seen as DIEC procedures work only under the IID assumption and thus fail to be operational in the sense defined above[5]. This includes tests concerned with the demonstration of entanglement via an entanglement witness (i.e. answering a yes–no question) [GT09, BBS+13, Ban14], as well as more quantitative analyses of entanglement measures such as the negativity and dimension witness [MBL+13]. In all of these works, the focus is on relating properties of a single multipartite state to some asymptotic statistics. An application of these works in an experimental setting is then most straightforwardly obtained under the assumption that the experiment consists in a repetition of identical rounds in which the same state is produced independently each time, i.e. under the IID assumption. The same assumption is inherent to claims regarding the amount of remaining entanglement which was not consumed by the estimation procedure: such claims rely on the assumption that more states, identical to those which were used for the testing phase, can be created by the source.

Another line of recent works deals with self-testing of high dimensional entangled states [McK16, CRSV16, Col17, CN16, NV17, CS17]. The goal of such works is more ambitious than entanglement certification; they aim to quantify the distance of the state used to perform the relevant tests from some specific target state, e.g. $|\Phi^L\rangle$ (up to local isometries, which cannot be excluded in the DI setting). Of course, once such a bound is derived it also implies lower bounds on various (continuous) entanglement measures.

A crucial disadvantage of these works is that they are not noise-tolerant in any realistic sense and therefore are not adequate when dealing with noisy sources of entanglement. For example, IID noisy sources, which produce many IID copies of a noisy entangled state, e.g. the two-qubit Werner state $\sigma = (1 - \xi)|\phi^+\rangle\langle\phi^+| + \xi\mathbb{I}/4$ for some constant (i.e. independent of the number of copies being created) noise value $\xi \in [0, 1]$, pass the considered self-tests only with negligible probability. Thus, no conclusion regarding the amount of entanglement produced by such IID noisy

[4] 'Malicious devices' should be understood here as devices that 'try to convince' a verifier that they produce highly entangled states while this is not the case. This is relevant, e.g. in the context of benchmarking, where one may argue, intentionally or not, that a given source is better than another even though this is not true.

[5] This statement also holds for works in the semi-DI setting; see, e.g. [CS16, MPB+16].

sources can be derived[6]. Furthermore, here as well, most self-testing results can only be used to describe the entanglement which was already consumed in the self-test. As far as we are aware, the only self-testing works where this is not the case are [CRSV16, NV17].

1.2. Distillable entanglement

There are many different ways of quantifying entanglement, some of which were already mentioned above (for surveys see [HHHH09, PV14]). One of the most basic and meaningful measures of entanglement is the *distillable entanglement*. Roughly speaking, given a multipartite state, its distillable entanglement describes the number of maximally entangled states that can be 'extracted' out of it by employing only local operations and classical communication (LOCC). Since LOCC cannot be used to create entanglement between the parties, such a process, termed entanglement distillation, indeed quantifies the entanglement of the initial state itself.

The task of entanglement distillation was first considered in [BBP+96, BBPS96, BDSW96]. There, two parties, Alice and Bob, share n copies of a bipartite mixed state σ. Their goal is to apply some LOCC to create $r < n$ copies of, say, $|\Phi^+\rangle$. This motivates the following, widely used, definition of distillable entanglement:

Definition 1 (Asymptotic IID distillable entanglement). The asymptotic IID distillable entanglement of a bipartite state $\sigma \in \mathcal{H}_A \otimes \mathcal{H}_B$ is given by

$$E_D^{\infty}(\sigma) = \sup\left\{ r \;\middle|\; \lim_{n\to\infty}\left(\sup_{\Gamma} F(\Gamma(\sigma^{\otimes n}), \Phi^{2^{rn}}) = 1\right)\right\}, \tag{1}$$

where Γ is an LOCC map (with respect to the bipartition of σ) and F is the fidelity.

Equation (1) describes a scenario in which one starts with n independent copies of σ and requires that, as n goes to infinity, the error of the distillation protocol goes to zero. This explains why we call it here the *asymptotic IID* distillable entanglement and not simply the distillable entanglement as it is usually called in the literature. However, as we claimed above, the sources that we consider do not necessarily produce IID states $\sigma^{\otimes n}$ (and they, definitely, do not emit $n \to \infty$ entangled states). Hence, $E_D^{\infty}(\sigma)$ does not truly quantify the entanglement produced in our scenario.

For our purpose, a more suitable entanglement measure is the *one-shot distillable entanglement* [BD10]. In the one-shot scenario Alice and Bob share a single copy of a bipartite state ρ and their goal is to convert it to the maximally entangled state $|\Phi^L\rangle$, for the maximal possible value L, using only LOCC. We say that the distillation protocol is successful when the resulting state is ε-close to $|\Phi^L\rangle$ for some fixed ε.

To state the definition of the one-shot distillable entanglement in a way comparable to its asymptotic IID counterpart E_D^{∞}, we consider $\rho \in \mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$ for a fixed n and some Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ (while ρ itself does not necessarily have the IID form $\sigma^{\otimes n}$). We then identify the one-shot distillation rate as $r = \log(L)/n$[7]. We can now use the following definition:

Definition 2 (One-shot distillable entanglement). Let $n \in \mathbb{N}_+$ and $\varepsilon \in [0, 1]$. The one-shot distillable entanglement of a bipartite state $\rho \in \mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$ is given by

$$E_D^{n,\varepsilon}(\rho) = \sup\left\{ \log(L)/n \;\middle|\; \sup_{\Gamma} F(\Gamma(\rho), \Phi^L) = 1 - \varepsilon \right\}, \tag{2}$$

where Γ is an LOCC map (with respect to the bipartition of ρ) and F is the fidelity.

$E_D^{n,\varepsilon}(\rho)$ describes the number of maximally entangled states which can be extracted, using LOCC, from a *single* copy of an arbitrary bipartite state ρ while allowing for some error ε. Hence, it captures the amount of entanglement available in ρ when using it as a resource in quantum information processing tasks.

1.2.1. Structure of the paper

The following sections are arranged as follows. In section 2 we present our contribution: our definition of an operational DIEC, the considered setting, and our results (the protocol and achieved rates). One can find all the necessary preliminary information and notation in section 3. Section 4 is devoted to presenting the main steps of

[6] Note that this issue is inherent to the distance measures used in all self-testing works (i.e. their objective) rather than some non-optimal properties of the specific tests considered in the mentioned works.

[7] For a given ρ there are different ways of choosing n, $\mathcal{H}_A$ and $\mathcal{H}_B$ such that $\rho \in \mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$. Thus, different choices can lead to different distillation *rates* $r = \log(L)/n$ (while $n\, E_D^{n,\varepsilon}(\rho)$ is independent of these choices). Later on there will be no ambiguity regarding the value of n and hence the used definition will be the one most relevant for us.

the proof and section 5 includes several remaining open questions. All the technical details of the proofs can be found in the appendix.

2. Our contribution: operational DIEC

After setting the stage in the previous section, we are now ready to state the objective of the current work—DI certification of one-shot distillable entanglement—and our results. We start in section 2.1 by introducing and motivating our definition of an operational DIEC protocol. We then explain in section 2.2 the exact setting of the source and measurement devices considered in our work. In section 2.3 we present our results.

2.1. The goal

Let us start by defining explicitly what we mean by a DIEC protocol and, by this, set the goal of this work. Given an uncharacterised source of entanglement producing n bipartite systems globally described by the state $\phi \in \mathcal{H}_{\tilde{A}}^{\otimes n} \otimes \mathcal{H}_{\tilde{B}}^{\otimes n}$ for some (unknown) Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ and (at least) two measurement devices not entangled to one another but otherwise uncharacterised, our goal is to find a DIEC protocol, employing only LOCC, that certifies that ϕ is highly entangled in a meaningful operational way. The certification protocol is going to act on ϕ, using the measurement devices, and we would like to claim (roughly) that, if the protocol does not abort, then the *final state* has high amount of one-shot distillable entanglement.

More precisely, we define a DIEC as follows.

Definition 3 (DIEC protocol). For any $n \in \mathbb{N}_+$ let $\varepsilon_{\text{dist}}$, ε_{snd}, $\varepsilon_{\text{cmp}} \in [0, 1]$ be the desired error constants, and $r \in [0, 1]$ the desired threshold distillation rate. Furthermore, let $\mathcal{S}^{\text{honest}}$ be a set of states ϕ^{honest} produced by a desired 'honest source' and $\mathcal{D}^{\text{honest}}$ the set of the desired 'honest measurement devices'[8]. Let P be a protocol *employing only* LOCC that given a state $\phi \in \mathcal{H}_{\tilde{A}}^{\otimes n} \otimes \mathcal{H}_{\tilde{B}}^{\otimes n}$ creates a state $\rho \in \mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n}$. We denote the final state conditioned on not aborting the protocol by $\rho_{|\Omega}$. The protocol P is said to be a DIEC protocol if the following two properties hold.

1. Noise-tolerance (completeness): The probability that P aborts when applied on any $\phi^{\text{honest}} \in \mathcal{S}^{\text{honest}}$ using honest measurement devices from $\mathcal{D}^{\text{honest}}$ is at most ε_{cmp}.
2. Entanglement certification (soundness): For any source and measurement devices either P aborts with probability greater than $1 - \varepsilon_{\text{snd}}$ when applied on ϕ or $E_D^{n,\varepsilon_{\text{dist}}}(\rho_{|\Omega}) \geqslant r$.

There are several important remarks to make regarding the above definition.

1. One possible example for an honest source is a source producing n independent copies of the Werner state $\sigma = (1 - \xi)|\phi^+\rangle\langle\phi^+| + \xi\mathbb{I}/4$ for some maximal value of $\xi > 0$. We can then define

$$\mathcal{S}^{\text{honest}} = \{((1 - \tilde{\xi})|\phi^+\rangle\langle\phi^+| + \tilde{\xi}\mathbb{I}/4)^{\otimes n} \mid \tilde{\xi} \leqslant \xi \}.$$

   $\mathcal{D}^{\text{honest}}$ can be defined to include, e.g. the measurement devices that apply the optimal measurements performed in the CHSH game. The noise-tolerance property, also termed completeness, states that the protocol should not abort, with high probability, for any $\phi^{\text{honest}} \in \mathcal{S}^{\text{honest}}$ even though some noise $\xi > 0$ is present. This implies (in combination with the soundness property) that P is able to certify the entanglement produced by the honest source. Of course, the honest sets $\mathcal{S}^{\text{honest}}$ and $\mathcal{D}^{\text{honest}}$ can be chosen in any way one wishes and depending on the experimental setting one has in mind. For instance, it could include states with a different noise model than the Werner state. In most cases the manufacture of the entanglement source (the experimentalist) has some 'guess' for a realistic description of the source and measurement devices. In most applications, these define the sets that should be chosen as the honest sets.

2. The entanglement certification, or soundness, property means that for *any* state produced by the unknown source and *any* measurement devices either the protocol identifies that the apparatus is not sufficiently good, and therefore the protocol aborts, or the post-protocol state $\rho_{|\Omega}$ is highly entangled, in the sense that there exists a way to distill (close to) nr maximally entangled states out of it. The protocol therefore certifies that $\rho_{|\Omega}$ is indeed useful for subsequent applications[9].

[8] Mathematically, this set can be defined as, e.g. a set of projectors applied by the measurement devices for the different possible inputs.

[9] Note that below we only claim that a distillation protocol *exists* but we do not present it explicitly.

3. To be able to quantify the entanglement produced by the *source itself*, rather than the entanglement that could be present inside the measurement devices, we need to assume that the measurement devices were not entangled before the start of the protocol. This situation is somehow similar to the requirement for measurement independence in a Bell test: if the measurement settings used in a Bell test are chosen by a device correlated with the source, then quantum statistics can be reproduced by local models [Hal11, BG11]. In practice, measurement independence is guaranteed by assuming that the devices used to produce the measurement settings behave independently from the rest of the setup. Similarly, we could assume that the measurement device of each party is independent from all other devices involved in the protocol. However, this hypothesis is stronger than necessary. We thus make the lighter assumption that the measurement devices share no entanglement with each other at the beginning of the protocol.

4. Only LOCC protocols should be considered as DIEC protocols since we would like to certify the distillable entanglement produced by the source and not the entanglement that may be produced by a protocol employing operations that cannot be explained via LOCC[10].

5. In a way, the definition of a DIEC can be seen as an extension of the so called SWAP technique [BNS+15], used in self-testing works, to the non-IID setting. Roughly speaking, by constructing the SWAP operators one can claim that a state close to, e.g. the maximally entangled state, up to local isometries can be 'extracted' out of the uncharacterised devices. The construction of a DIEC protocol implies that one can extract a state close to *many* maximally entangled states using LOCC.

6. The above definition allows us to use DIEC protocols as a way to *compare* different entanglement sources which are otherwise incomparable. For example, one can consider two sources developed by different experimental groups. One source produces, say, many identical copies of the Werner state while the other creates many identical copies of perfect partially entangled states. Each group is free to choose a DIEC protocol, as in definition 3, that will result in the highest lower bound on the distillable entanglement produced by its source. If one group wishes to claim that its source is 'better', then the certified distillation rate of their source (which can be verified by any user) must be higher than that of the competing source. The protocols allow us to *benchmark* one source against the other in a meaningful way by allowing any user to verify a claimed lower bound on the produced entanglement available for further applications.

2.2. The setting: source and measurement devices

Here we describe the theoretical setting considered in our work to which our results apply, i.e. under which we will prove the soundness of our DIEC protocol. Different variants of this setting can be chosen depending on one's interest; we choose the presented one as we believe it is both realistic and simple to discuss. This setting is compatible with the standard assumptions used in the DI setting (e.g. when testing the CHSH Bell inequality in DI quantum key distribution or randomness generation protocols). In particular, we employ the following standard assumptions: the measurement devices are separated in space and are the verifier can restrict their communication in different stages of the protocol and the verifier holds a trusted random number generator and is able to make basic classical calculations required to run the protocol. In addition, we assume that quantum physics is correct.

2.2.1. The theoretical setting: source, measurement devices, and quantum registers

In order to be able to talk about the entanglement available *after* running the DIEC protocol we must be able to have a well defined state at hand, whose entanglement we are quantifying. For this, we consider a theoretical setting which fulfils the following two conditions.

First, the production of any entangled states (if such are being created) can be attributed only to the source. In other words, we assume that the measurement devices *neither produce nor hold* additional entanglement. This implies a distinction between the source and the measurement devices.

Second, the state produced by the source, or the 'post-protocol state', can be kept in some registers, i.e. quantum memory. This is necessary at the theoretical level since we wish to discuss the remaining entanglement in a meaningful way. The registers are *trusted*, in the sense that they cannot be manipulated by the devices after they are accessed during the protocol. Indeed, if the devices are allowed to, e.g. measure the registers in which the final state is being kept then, clearly, one cannot say anything regarding the entanglement left in the system[11].

[10] Another option may be to consider protocols that employ only separability preserving operations; see [Rai97, CDKL01, BD11a].

[11] This also fits the distinction we made above between the source and the measurement devices. If the entanglement is produced by the measurement devices and is kept 'inside of them' then they can also destroy whatever is left in the end of the protocol. While the results are closely related, Lemma 5 does not follow from [PAB+09].

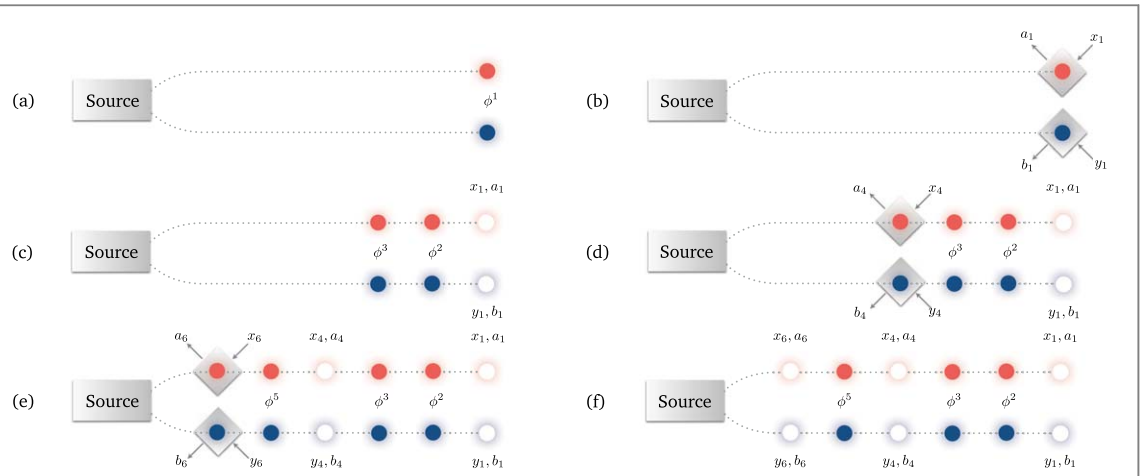

Figure 1. Illustration of the considered setting. (a) The source emits the first state in the sequence ϕ^1. (b) The verifier chooses to measure the first state. The two measurement devices, denoted by the grey squares, measure ϕ^1 according to the inputs x_1, y_1 given by the verifier. They output a_1, b_1. (c) The next states ϕ^2, ϕ^3 are produced by the source one after the other. They can depend on the values x_1, y_1, a_1, b_1. (d) The source emits ϕ^4 and the verifier chooses to measure it using his measurement devices. The actions of the measurement devices can depend on the classical information of the previous rounds: the values x_1, y_1, a_1, b_1 as well as the knowledge that ϕ^2, ϕ^3 were not measured. (e) The next states are produced by the source. The verifier chooses to measure only ϕ^6. (f) The final state which is kept in the memory.

Note that by considering the one-shot distillable entanglement we are already hinting that one should be able to apply an entanglement distillation protocol on the certified state (at least in theory). To apply such a protocol, the state must be available somewhere so it can be manipulated. Defining some quantum registers in which the state is being kept is therefore necessary in our context. The quantum registers are merely a theoretical tool, they are not needed in an experimental implementation of our DIEC protocol.

The reminder of the current section is devoted to explaining precisely the setting, fulfilling the above conditions that we consider in the current work. The setting is illustrated in figure 1.

The entity wanting to certify the entanglement produced by the source is called the *verifier*. To run the DIEC protocol the verifier holds two separated (space-like or otherwise shielded) measurement devices. As we consider two measurement devices we can also treat the verifier as two parties, Alice and Bob (both the 'verifier' and 'Alice and Bob' are used below, depending on the context).

We consider an uncharacterised source emitting a sequence of entangled quantum systems. (For example, one can imagine a source that emits pairs of entangled photons one after the other.) We denote by $n \in \mathbb{N}_+$ the number of systems produced by the source, i.e. the number of times the source is being used during the DIEC protocol. For every $i \in [n]$, the source produces some unknown state ϕ^i that is being kept in the memory. The verifier then chooses whether to measure the state using his measurement devices or not. If he measures the state then the classical inputs given to the devices and the outputs produced by them are kept in a classical memory. If ϕ^i is not measured then it is kept as is in a quantum memory.

The next state produced by the source, ϕ^{i+1}, can depend on all of the previous classical information, i.e. which of the previous states were measured and what were the inputs and outputs in those rounds. Note that while ϕ^{i+1} can depend on whether a test was made in a previous round $j < i + 1$, the source and the measurement devices have no further access to the quantum states ϕ^j which *were not measured*. That is, we assume that each ϕ^j which was not measured is not affected by the devices after the jth round. Again, this is necessary as otherwise the remaining entanglement can be destroyed by the devices. We emphasise that this *does not imply* that the systems of the different rounds cannot be entangled with one another.

We use the above model in our protocol and its analysis. However, we remark that the protocol and analysis can be adapted to capture other models with a sequential structure.

2.3. Results

2.3.1. Our DIEC protocol and the achieved rates

Our DIEC protocol is presented as protocol 1. It is based on the CHSH inequality (see section 3.2 for the necessary basic information). Similar protocols can also be considered for other Bell inequalities.

Protocol 1. DIEC protocol (based on the CHSH inequality)

Arguments:

D—untrusted measurement device of two components with inputs and outputs set $\{0, 1\}$.
$n \in \mathbb{N}_+$—number of rounds.
γ—the probability of a test.
ω_{exp}—expected winning probability in the CHSH game.
$\delta_{\text{est}} \in (0, 1)$—width of the statistical confidence interval for the estimation test.

For every round $i \in [n]$ do Steps 2–10:
  Let ϕ^i denote the bipartite state produced by the source in this round.
  Set $A_i, B_i, X_i, Y_i, W_i = \perp$.
  Choose $T_i = 1$ with probability γ and $T_i = 0$ otherwise.
  If $T_i = 1$:
    Choose the inputs $X_i, Y_i \in \{0, 1\}$ uniformly at random.
    Measure ϕ^i using D with the inputs X_i, Y_i and record the outputs as $A_i, B_i \in \{0, 1\}$.
    Set $W_i = 1$ if the CHSH game is won and $W_i = 0$ otherwise.
  If $T_i = 0$:
    Keep ϕ^i in the registers $\hat{A}_i \hat{B}_i$.
Abort if $\sum_i \chi(T_i = 1) W_i < (\omega_{\text{exp}} \gamma - \delta_{\text{est}}) n$.

As mentioned in the previous section, we consider a source which produces a sequence of bipartite states, denoted by ϕ^i for $i \in [n]$. The verifier chooses whether to measure ϕ^i or keep it as is in the memory. The register T_i describes whether a test was performed or not. If a test is performed, the registers $X_i Y_i A_i B_i$ hold the classical inputs and outputs. The register W_i is set to 1 when the CHSH game is won in the ith round and 0 otherwise. When a test is not being performed, the state ϕ^i is kept in the quantum registers $\hat{A}_i \hat{B}_i$. We allow the source to 'know' all the classical information of the previous rounds, i.e. $(TABXYW)_{1,\ldots,i-1}$ and hence ϕ^i can also depend on this information.

We denote the state in the end of protocol 1, before Step 11, by

$$\rho = \rho_{\hat{A}\hat{B}ABXYTW}.$$

Denote by χ the indicator function, i.e. $\chi = 1$ if $T_i = 1$, and $\chi = 0$ otherwise. Let Ω be the event $\sum_i \chi(T_i = 1) W_i \geqslant (\omega_{\text{exp}} \gamma - \delta_{\text{est}}) n$, i.e. the event of not aborting the protocol. We denote the state after the end of the protocol, conditioned on not aborting, by

$$\rho_{|\Omega} = \rho_{\hat{A}\hat{B}ABXYTW|\Omega}.$$

We define the bipartition of ρ and $\rho_{|\Omega}$ as $Q^A = \hat{A}$ versus $Q^B = \hat{B}ABXYTW$. It is easy to see that for this bipartition protocol 1 employs only LOCC for *any* measurement devices in our setting (i.e. initially non-entangled devices that do not communicate during the time of the measurement).

Our main result states that protocol 1 is indeed an operational DIEC protocol. That is, it fulfils the requirements of definition 3.

As the honest source and devices we choose to consider a source that produces identical and independent copies of a state $\phi^i = \sigma$ and measurement devices that apply the same measurements in each round when they are used. The state σ and the measurements are such that the winning probability achieved in the CHSH game is at least ω_{exp}. For example, one can choose $\mathcal{D}^{\text{honest}}$ to include the measurement devices that apply the optimal measurements performed in the CHSH game and the set $\mathcal{S}^{\text{honest}}$ to include all states $\phi^{\text{honest}} = \sigma^{\otimes n}$ for σ any noisy maximally entangled state that will result in winning probability $\geqslant \omega_{\text{exp}}$.

The following theorem states our main result.

Theorem 4 (Main theorem). *For any $n \in \mathbb{N}_+$, $\varepsilon_{\text{dist}}, \varepsilon_{\text{snd}} \in [0, 1]$, and $\varepsilon_{\text{smo}} \in [0, \sqrt{\varepsilon_{\text{dist}}})$, protocol 1 is a DIEC protocol with the following.*

1. *Noise-tolerance (completeness): The probability that* P *aborts when applied on any $\phi^{\text{honest}} \in \mathcal{S}^{\text{honest}}$ using honest measurement devices from $\mathcal{D}^{\text{honest}}$ is at most $\varepsilon_{\text{cmp}} \leqslant \exp(-2n\delta_{\text{est}}^2)$.*
2. *Entanglement certification (soundness): For* any *source and measurement devices in the considered setting, either protocol 1 aborts with probability greater than $1 - \varepsilon_{\text{snd}}$ when applied on ϕ or $E_D^{n,\varepsilon_{\text{dist}}}(\rho_{|\Omega}) \geqslant r$ for $r = \log(L)/n$ and*

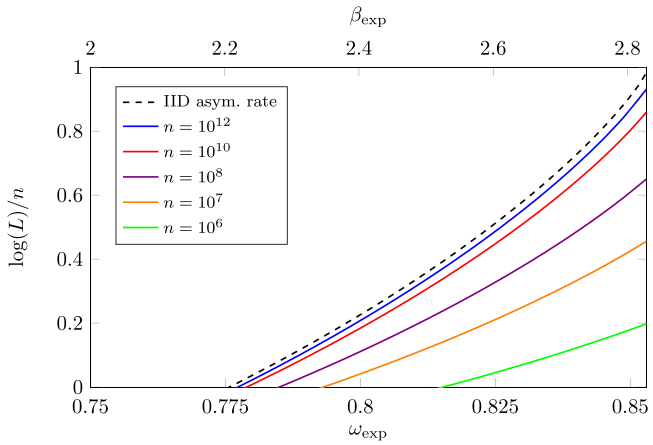

Figure 2. The one-shot distillable entanglement rate $r = \log L/n$ as a function of the expected winning probability in the CHSH game ω_{exp} (or β_{exp}) for different values of n (the analytical form is given in equation (3)). The error parameters were chosen to be $\varepsilon_{\text{dist}} = \varepsilon_{\text{snd}} = 10^{-5}$ and the completeness parameter was set to $\varepsilon_{\text{cmp}} = 10^{-2}$. All other values were chosen so that the distillable entanglement rate is maximised. The dashed curve describes the distillable entanglement rate which can be certified in the IID asymptotic case.

$$\log L = -n \cdot \eta_{\text{opt}}(\varepsilon_{\text{smo}}, \varepsilon_{\text{snd}}) - 4\log\left(\frac{1}{\sqrt{\varepsilon_{\text{dist}}} - \varepsilon_{\text{smo}}}\right) \in \Omega(n), \tag{3}$$

where η_{opt} is defined in equation (14).

The exact bound on the certified distillation rate r is not very informative, so we postpone discussing its explicit form to later. Instead, we plot the distillation rate in figure 2 for some choices of parameters. Clearly, the optimal distillation rate is upper bounded by 1 and hence $\log L \in \Omega(n)$, as achieved in our work, is optimal in terms of the asymptotic dependency on n.

To discuss the tightness of the derived rates more concretely we need to first say few words about our proof technique. To prove theorem 4 we use two independent results. The first is a lower bound on the one-shot distillable entanglement in terms of the negative of the *conditional smooth max-entropy* $H_{\max}^{\varepsilon_{\text{smo}}}(Q^A|Q^B)_{\rho_{|\Omega}}$ [WTB17]. This allows us to reduce the task of proving theorem 4 to that of (upper) bounding $H_{\max}^{\varepsilon_{\text{smo}}}(Q^A|Q^B)_{\rho_{|\Omega}}$. The main tool used to derive a bound on the smooth max-entropy is the entropy accumulation theorem (EAT) [DFR16].

Our bound on $H_{\max}^{\varepsilon_{\text{smo}}}(Q^A|Q^B)_{\rho_{|\Omega}}$ is tight to first order in n and hence cannot be significantly improved. In particular, this means that the regime of Bell violations $\omega_{\text{exp}} \lessapprox 0.775$ in which our distillation rate is zero, although the verifier observes a violation, cannot be improved by better bounding $H_{\max}^{\varepsilon_{\text{smo}}}(Q^A|Q^B)_{\rho_{|\Omega}}$.

Qualitatively, it is not surprising that such a regime exists. Indeed, it was already shown that, in general, Bell non-locality is fundamentally different from distillable entanglement (this stands in contrast to the so called Peres conjecture). That is, there are *bound entangled states* (i.e. entangled state which cannot be distilled) that can be used to violate some Bell inequalities [VB14]. For the CHSH inequality, however, this is not the case [Mas06]: bound entangled states cannot be used to violate the CHSH inequality. Hence, *asymptotically*, one should be able to certify distillable entanglement for *any* violation $\omega_{\text{exp}} > 0.75$.

This implies that, although our bound on $H_{\max}^{\varepsilon_{\text{smo}}}(Q^A|Q^B)_{\rho_{|\Omega}}$ is tight, the relation between the smooth max-entropy and the one-shot distillable entanglement is not; it only gives a *lower bound* on the amount of distillable entanglement but it should be possible to distill more (see section 5 for more details).

2.3.2. von Neumann entropy as a function of the CHSH violation

As a step in our proof of theorem 4 we derive an upper bound on the conditional von Neumann entropy $H(\hat{A}_i|\hat{B}_i)_\sigma$ for any Bell-diagonal quantum state $\sigma_{\hat{A}_i\hat{B}_i}$ that can be used to win the CHSH game with winning probability $\omega \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$. The conditional von Neumann entropy is negative only when evaluated over entangled states. Thus, an upper bound on it can be seen, by itself, as a quantitive certificate of entanglement (though not in our operational sense). Such a bound might be of independent interest in other contexts.

Inspired by [PAB+09], in which a 'dual quantity'[12] was bounded, we derive a bound for Bell diagonal states. We prove the following:

[12] In [PAB+09] a bound on $H(A|E)$ was derived, where A is the measurement outcome on Alice's side (when measuring $\sigma_{\hat{A}_i}$ in our notation) and E describes a system used to purify $\sigma_{\hat{A}_i\hat{B}_i}$.

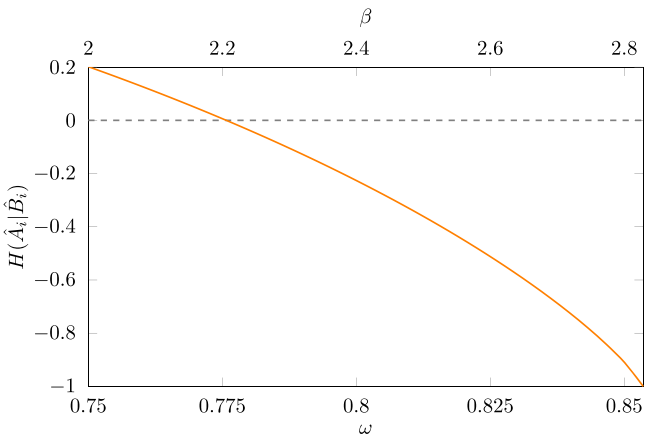

Figure 3. The upper bound on $H(\hat{A}_i|\hat{B}_i)$ as a function of the CHSH winning probability ω or the violation β.

Lemma 5. *For any Bell diagonal state $\sigma_{\hat{A}_i\hat{B}_i}$ that can be used to violate the CHSH inequality with violation $\omega \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$,*

$$H(\hat{A}_i|\hat{B}_i)_\sigma \leqslant 2h\left(\frac{1}{2} - \frac{2\omega - 1}{\sqrt{2}}\right) - 1, \tag{4}$$

where h is the binary entropy function.

The bound given in equation (4) is plotted in figure 3. The bound is tight, i.e. there exist states that saturate it. As mentioned above, the 'interesting' regime of the bound is that in which the conditional entropy is negative, i.e. $\omega \gtrapprox 0.775$ ($\beta \gtrapprox 2.2$). A similar bound, but for the conditional max-entropy $H_{\max}(\hat{A}_i|\hat{B}_i)_\sigma$ was derived in (AFRV98, supplementary note 2). As $H_{\max}(\hat{A}_i|\hat{B}_i)_\sigma \geqslant H(\hat{A}_i|\hat{B}_i)_\sigma$, their bound can also be used to upper bound the von Neumann entropy. However, using our bound directly leads to better quantitive results. In particular, the bound derived in [AFRV98] only leads to a negative upper bound for $\beta \gtrapprox 2.5$.

The observation that there exists a regime in which the conditional entropy is positive although the CHSH inequality is violated is not new. Indeed, it was already known that that some states, e.g. the Werner state, can be used to violate the CHSH inequality while presenting positive conditional entropy [FBB17].

The fact that the bound (4) on the conditional entropy is negative as soon as the winning probability is larger than $\omega \gtrapprox 0.775$ allows our scheme to certify entanglement for honest implementations based only on their CHSH winning probability. In particular, the honest implementation might be different than the Werner state we considered above as an example. At the same time, since the minimum winning probability required to obtain a useful bound on the conditional entropy is larger than 0.75, a significant Bell violation is required. This has implications on the critical detection efficiency of the scheme: whereas a violation of the CHSH Bell inequality is possible as soon as the detection efficiency is larger than 66.7% [Ebe93], the minimal detection efficiency required to reach a CHSH value of 2.2 is 85.3%.

3. Preliminaries

3.1. General notation

The set $\{1, 2, \ldots, n\}$ is denoted by $[n]$.

All logarithms are in base 2. Random variables (RV) are denoted by capital letters while specific values are denoted by small letters. Sets are denoted with calligraphic fonts. For example, we use X_i to denote a RV over $\mathcal{X}$ and x_i to denote a certain value $x_i \in \mathcal{X}$.

Most RV will refer to a specific $i \in [n]$ denoted in their subscript, as in X_i above. RV describing a range between $i, j \in [n]$ for $i < j$ are written as $X_{i,\ldots,j} = X_i, \cdots, X_j$. When no subscript appears then the range is from 1 to n, i.e. $X = X_1, \cdots, X_n$.

The fidelity of two quantum states is given by $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$.

3.2. The CHSH inequality and game

In a bipartite scenario where each party, Alice and Bob, can perform one of two possible measurements, indexed by X_i, $Y_i \in \{0, 1\}$, and with outcomes denoted by A_i, $B_i \in \{0, 1\}$, a Bell inequality is a linear constraint on the

conditional probability distributions $P(A_i, B_i|X_i, Y_i)$ which is satisfied by all local hidden variable models [BCP+14]. One Bell inequality of special interest is the so-called Clauser–Horne–Shimony–Holt (CHSH) inequality [CHSH69], which takes the form

$$\beta = \sum_{a_i,b_i,x_i,y_i} (-1)^{a_i+b_i+x_iy_i} P(a_i, b_i|x_i, y_i) \leqslant 2.$$

This inequality admits a maximum quantum violation of $\beta = 2\sqrt{2}$. To achieve this maximal violation Alice and Bob can share the maximally entangled state $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and measure it with the following measurements: Alice's measurements $X_i = 0$ and $X_i = 1$ correspond to the Pauli operators σ_z and σ_x respectively and Bob's measurements $Y_i = 0$ and $Y_i = 1$ to $(\sigma_z + \sigma_x)/\sqrt{2}$ and $(\sigma_z - \sigma_x)/\sqrt{2}$ respectively. We can therefore restrict our attention to $\beta \in [2, 2\sqrt{2}]$.

The CHSH inequality can be equivalently expressed in terms of a game. The game is defined via the winning condition $w_{\text{CHSH}} = 1$ if $A_i \oplus B_i = X_i \cdot Y_i$. The optimal quantum strategy described above achieves winning probability $\omega = \frac{2+\sqrt{2}}{4} \approx 0.85$, while the optimal classical strategy achieves a winning probability of 0.75. The relation between the winning probability in the game and the violation of CHSH inequality is given by $\omega = 1/2 + \beta/8$ and we have $\omega \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$.

3.3. Quantum entropies and Markov chains

3.3.1. Shannon entropy

The Shannon entropy of a probability distribution $\{p_i\}$ is given by $H(\{p_i\}) = -\sum_{p_i} p_i \log(p_i)$. For a Bernoulli process with probability p the we use the binary entropy function $h(p) = -p\log(p) - (1-p)\log(1-p)$.

3.3.2. von Neumann entropy

The von Neumann entropy $H(\rho)$ of a quantum state ρ is given by $H(\rho) = -\text{Tr}(\rho \log \rho)$. Given a bipartite state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ the conditional von Neumann entropy is

$$H(A|B)_{\rho_{AB}} = H(\rho_{AB}) - H(\rho_B).$$

When the state on which the entropy is evaluated is clear from the context we drop the subscript and write $H(A|B)$.

3.3.3. Max-entropy

The explicit definition of the quantum max-entropy will not be of use in the current work. Yet, we state it here for completeness. The conditional smooth max-entropy of a bipartite quantum state ρ_{AB} is given by

$$H^{\varepsilon}_{\max}(A|B)_{\rho_{AB}} = \log \inf_{\sigma_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})} \sup_{\tau_B} \|\sigma_{AB}^{\frac{1}{2}} \tau_B^{-\frac{1}{2}}\|_1^2,$$

where $\mathcal{B}^{\varepsilon}(\rho_{AB})$ the set of sub-normalised states σ_{AB} with $P(\rho_{AB}, \sigma_{AB}) \leqslant \varepsilon$ and P is the purified distance [TCR10].

3.3.4. Markov chains

A tripartite quantum state ρ_{ABC} is said to fulfil the Markov chain condition $A \leftrightarrow B \leftrightarrow C$ if $I(A\colon C|B) = 0$, where $I(A\colon C|B) = H(AB) + H(BC) - H(B) - H(ABC)$ is the conditional mutual information. Note that $I(A\colon C|B) = 0$ if and only if given B, A and C are independent.

3.4. The EAT

The EAT ([DFR16], theorem 4.4) gives us a way of bounding the amount of smooth min- or max-entropy accumulated during a sequential process fulfilling certain conditions. In contrast to previous works where a bound on the smooth min-entropy was derived using the EAT, the current work uses the EAT to bound the smooth max-entropy. We state here the necessary details in the context of our work.

To apply the EAT one needs to define 'EAT channels' which describe the sequential process under consideration (in our case, for example, the channels are defined via the actions of our DIEC protocol). EAT channels are defined as follows.

Definition 6 (EAT channels). EAT channels $\mathcal{N}_i\colon R_{i-1} \to R_i O_i S_i W_i$, for $i \in [n]$, are CPTP maps such that, for all $i \in [n]$, as follows.

1. O_i are finite dimensional quantum systems of dimension d_{O_i} and W_i are finite-dimensional classical systems (RV). S_i and R_i are arbitrary quantum systems.

2. For any $i \in [n]$ and any input state $\sigma_{R_{i-1}}$, the output state $\sigma_{R_i O_i S_i W_i} = \mathcal{N}_i(\sigma_{R_{i-1}})$ has the property that the classical value W_i can be measured from the marginal $\sigma_{O_i S_i}$ without changing the state. That is, for the map $\mathcal{T}_i$: $O_i S_i \to O_i S_i W_i$ describing the process of deriving W_i from O_i and S_i, it holds that $\mathrm{Tr}_{W_i} \circ \mathcal{T}_i(\sigma_{O_i S_i}) = \sigma_{O_i S_i}$.

3. For any initial state $\tau^0_{R_0}$, the final state $\tau_{OSW} = (\mathrm{Tr}_{R_n} \circ \mathcal{N}_n \circ ... \circ \mathcal{N}_1)\, \tau^0_{R_0}$ fulfils the Markov chain condition $O_{1...i-1} \leftrightarrow S_{1...i-1} \leftrightarrow S_i$ for each $i \in [n]$.

We will use below the following notation. Given a value $w = w_1, \ldots, w_n \in \mathcal{W}^n$, where $\mathcal{W}$ is a finite alphabet, we denote by freq_w the probability distribution over $\mathcal{W}$ defined by $\mathrm{freq}_w(\tilde{w}) = \frac{|\{i \mid w_i = \tilde{w}\}|}{n}$ for $\tilde{w} \in \mathcal{W}$. If τ is a state classical on W we write $\Pr[w]_\tau$ to denote the probability that τ assigns to w.

Definition 7 (Max-tradeoff functions). Let $\mathcal{N}_1, \ldots, \mathcal{N}_n$ be a family of EAT channels. Let $\mathcal{W}$ denote the common alphabet of $W_1, \ldots, W_n$. A concave[13] function $f_{\max}$ from the set of probability distributions p over $\mathcal{W}$ to the real numbers is called a *max-tradeoff function* for $\{\mathcal{N}_i\}$ if it satisfies

$$f_{\max}(p) \geqslant \sup_{\sigma_{R_{i-1}}:\mathcal{N}_i(\sigma)_{W_i}=p} H(O_i|S_i)_{\mathcal{N}_i(\sigma)}$$

for all $i \in [n]$, where the supremum is taken over all input states of $\mathcal{N}_i$ for which the marginal on W_i of the output state is the probability distribution p.

The statement of the EAT, relevant for the smooth max-entropy, is given below.

Theorem 8 (EAT). *Let $\mathcal{N}_i$: $R_{i-1} \to R_i O_i S_i W_i$ for $i \in [n]$ be EAT channels as in definition 6, $\tau_{OSW} = (\mathrm{Tr}_{R_n} \circ \mathcal{N}_n \circ ... \circ \mathcal{N}_1)\tau_{R_0}$ be the final state, Ω an event defined over $\mathcal{W}^n$, $\Pr[\Omega]_\tau$ the probability of Ω in τ, and $\tau_{|\Omega}$ the final state conditioned on Ω. Let $\varepsilon_{\mathrm{smo}} \in (0, 1)$. For $f_{\max}$ a max-tradeoff function for $\{\mathcal{N}_i\}$ as in definition 7 and any $t \in \mathbb{R}$ such that $f_{\max}(\mathrm{freq}_w) \leqslant t$ for any $w \in \mathcal{W}^n$ for which $\Pr[w]_{\tau_{|\Omega}} > 0$,*

$$H^{\varepsilon_{\mathrm{smo}}}_{\max}(O|S)_{\tau_{|\Omega}} \leqslant nt + v\sqrt{n},$$

where $v = 2(\log(1 + 2d_{O_i}) + \lceil \|\nabla f_{\max}\|_\infty \rceil)\sqrt{1 - 2\log(\varepsilon_{\mathrm{smo}} \cdot \Pr[\Omega]_\tau)}$ and d_{O_i} denotes the dimension of $\mathrm{O_i}$.

We used above a slightly and trivially modified statement of the EAT, compared to that of [DFR16]. The definition of the max-tradeoff function used in [DFR16] consideres $H(O_i|S_i R')$ where R' is isomorphic to to R_{i-1}. For the calculation of the supremum one can always assume that the system on R' is in product with the rest of the system and hence drop it here (see Remark 4.2 in [DFR16]). Furthermore, in [DFR16] the EAT is stated with $H^{\varepsilon_{\mathrm{smo}}}_{\max}(O|SE)$ where E denotes a system extending the initial state τ_{R_0}, i.e. we have $\mathrm{Tr}_E(\tau_{R_0 E}) = \tau_{R_0}$. The E, in fact, can be dropped as the theorem must hold for any $\tau_{R_0 E}$ and, hence, in particular for $\tau_{R_0 E} = \tau_{R_0} \otimes \tau_E$ as in theorem 8, for which the conditional smooth max-entropy is maximal. In our context, an alternative way of thinking about this is to note that our goal is to bound $H^{\varepsilon_{\mathrm{smo}}}_{\max}(O|S)_{\tau_{|\Omega}}$. As it clearly depends only on the registers O and S, E does not take part in the proof and we can, w.l.o.g., consider an initial state of the form $\tau_{R_0} \otimes \tau_E$. The final state then also has a tensor product form.

4. Main parts of the proof

In this section we present the main steps and ideas used in the proof of theorem 4. The full details are given in the appendix.

4.1. Modified protocol

As explained above, our goal is to show that there exists an entanglement distillation protocol that can distill the entanglement present in $\rho_{|\Omega}$. Instead of considering an entanglement distillation protocol that acts on the state $\rho_{|\Omega}$ directly, we consider a slightly modified scenario. The modified scenario will result in a state $\tau_{|\Omega}$, closely related to $\rho_{|\Omega}$, from which at least the same amount of entanglement can be distilled.

Concretely, we modify the real DIEC, protocol 1, to define the modified protocol stated as protocol 2. The only difference between this protocol and protocol 1 is in Steps 11 and 12, which we explain below.

[13] Let $\hat{\Omega}$ be a set of frequencies defined via $\mathrm{freq}_w(\tilde{w}) \in \hat{\Omega}$ if and only if $\tilde{w} \in \Omega$. We can consider concave functions, in contrast to affine ones [DFR16], since the event Ω defined in the current work results in a *convex set* $\hat{\Omega}$.

We remark that protocol 2 is being used only as a step in the proof of our theorem. We do not claim at any point that this modified protocol can be implemented by the verifier given the uncharacterised devices (in fact, it cannot). It will become clear later on why the modified protocol is, nevertheless, needed in our proof.

Protocol 2. Modified DIEC (based on the CHSH inequality)

Arguments:

D—untrusted measurement device of two components with inputs and outputs set $\{0, 1\}$.

$n \in \mathbb{N}_+$—number of rounds.

ω_{exp}—expected winning probability in the CHSH game.

$\delta_{\text{est}} \in (0, 1)$—width of the statistical confidence interval for the estimation test.

1: For every round $i \in [n]$ do Steps 2–12:
2: Let ϕ^i denote the bipartite state produced by the source in this round.
3: Set $A_i, B_i, C_i, D_i, X_i, Y_i, W_i = \perp$.
4: Choose $T_i = 1$ with probability γ and $T_i = 0$ otherwise.
5: If $T_i = 1$:
6: Choose the inputs $X_i, Y_i \in \{0, 1\}$ uniformly at random.
7: Measure ϕ^i using D with the inputs X_i, Y_i. Record the outputs as $A_i, B_i \in \{0, 1\}$.
8: Set $W_i = 1$ if the CHSH game is won and $W_i = 0$ otherwise.
9: If $T_i = 0$:
10: Keep ϕ^i in the registers $\hat{A}_i\hat{B}_i$.
11: Apply the projection as in equation (6) to create $\bar{\rho}^{\,i}_{\hat{A}_i\hat{B}_iC_iD_i}$.
12: Choose a random unitary U uniformly within the set $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ and apply it to create $\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_iC_iD_i}$ as in equation (7).
13: Abort if $\sum_i \chi(T_i = 1) W_i < (\omega_{\text{exp}}\gamma - \delta_{\text{est}})n$.

4.1.1. First modification: reduction to qubits

When a test is not being performed by the verifier the state produced by the source is being kept in the memory. In protocol 1 the state is being kept as is. In contrast, in protocol 2 we first project the state such that the resulting state is a two qubit state. The projection, described by two local projections, can obviously only decrease the amount of entanglement of the kept state.

As one can guess, we define the specific projection via the measurements performed by the measurement devices[14]. Formally, due to Jordan's lemma (see, e.g. [Sca12], lemma 4.1), the measurements operators $\Pi^{A_i}_{a|x}$, $\Pi^{B_i}_{b|y}$ of each party, Alice and Bob, in each round $i \in [n]$ can be written in a suitable local basis as

$$\begin{aligned}
\Pi^{A_i}_{0|0} - \Pi^{A_i}_{1|0} &= \oplus_{c_i} A_0^{c_i} = \oplus_{c_i} \sigma_z^{c_i} \\
\Pi^{B_i}_{0|0} - \Pi^{B_i}_{1|0} &= \oplus_{d_i} B_0^{d_i} = \oplus_{d_i} \sigma_z^{d_i} \\
\Pi^{A_i}_{0|1} - \Pi^{A_i}_{1|1} &= \oplus_{c_i} A_1^{c_i} = \oplus_{c_i} \cos(\alpha_{c_i})\sigma_z^{c_i} + \sin(\alpha_{c_i})\sigma_x^{c_i} \\
\Pi^{B_i}_{0|1} - \Pi^{B_i}_{1|1} &= \oplus_{d_i} B_1^{d_i} = \oplus_{d_i} \cos(\beta_{d_i})\sigma_z^{d_i} + \sin(\beta_{d_i})\sigma_x^{d_i}
\end{aligned}$$

with parameters α_{c_i} and β_{d_i} and the different σ's denoting the Pauli matrices.

With this in mind, when $T_i = 0$ we let each party perform a measurement on $\rho^i_{\hat{A}_i\hat{B}_i}$ (which is identical to ϕ^i in the case $T_i = 0$) according to the following operators:

$$\begin{aligned}
\Pi^{\hat{A}_i}_{c_i} &= |c_i\rangle\langle c_i| \otimes \mathbb{I} \\
\Pi^{\hat{B}_i}_{d_i} &= |d_i\rangle\langle d_i| \otimes \mathbb{I}.
\end{aligned} \tag{5}$$

This gives each party the index of her/his respective Jordan blocks c_i and d_i. We denote by $p(c_i, d_i)$ the probability of observing the indices c_i, d_i. The global system is projected into the two-qubit state

$$\bar{\rho}^{\,i}_{\hat{A}_i\hat{B}_iC_iD_i} = \frac{(\Pi^{\hat{A}_i}_{c_i} \otimes \Pi^{\hat{B}_i}_{d_i})\rho^i_{\hat{A}_i\hat{B}_i}(\Pi^{\hat{A}_i}_{c_i} \otimes \Pi^{\hat{B}_i}_{d_i})}{\text{Tr}\,(\Pi^{\hat{A}_i}_{c_i} \otimes \Pi^{\hat{B}_i}_{d_i}\rho^i_{\hat{A}_i\hat{B}_i})} \otimes |c_i\rangle\langle c_i| \otimes |d_i\rangle\langle d_i|, \tag{6}$$

where we added a classical register for the measurement outcomes. At this stage, the parties exchange their indices c_i and d_i with each other. It is easy to see that the considered projection step can be described via LOCC.

Note the following.

[14] These are unknown to the verifier but, as mentioned above, protocol 2 does not need to be implemented in practice by the verifier. The only thing that matters is that such projectors *exist*.

1. Due to the projection, the registers $\hat{A}_i\hat{B}_i$ are now qubit registers.
2. Since the projectors given in equation (5) are defined via the measurements used by the device when $T_i = 1$, they could also be applied when $T_i = 1$ without changing the resulting state. This is made formal in lemma 9 below (for the proof see appendix A).

Lemma 9. *Consider a scenario in which the projection to the two qubit space is applied on the state ϕ^i directly after it is produced by the source in the* ith *round (i.e. before choosing the value of* $\mathrm{T_i}$*). Denote the resulting state in the end of the* ith *round in such a case by $\bar{\bar{\rho}}^i$. Then*

$$\bar{\bar{\rho}}^{\,i}_{\hat{A}_i\hat{B}_iA_iB_iC_iD_iX_iY_i} = \bar{\rho}^{\,i}_{\hat{A}_i\hat{B}_iA_iB_iC_iD_iX_iY_i},$$

where $\bar{\rho}^i$ is as defined in equation (6).

It is clear that for the state $\bar{\bar{\rho}}^i$ the measurements in the case of $T_i = 1$ act on the subspace of the same Hilbert space which is otherwise kept in $\hat{A}_i\hat{B}_i$ when $T_i = 0$, since the projection is made before choosing the value of T_i. Lemma 9 implies that this is also the case when the projection is applied only when $T_i = 0$ as in our modification. Without this property one could not argue that the test rounds, i.e. those for which $T_i = 1$, represent also the rounds with $T_i = 0$.

4.1.2. Second modification: reduction to Bell diagonal states

After the projection, we apply another step that can be seen as a symmetrisation step, also called twirling in the literature. In each round for which $T_i = 0$, a one-qubit unitary U is chosen uniformly within the set $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ and applied on both systems $\hat{A}_i$ and $\hat{B}_i$. That is,

$$\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_i|c_i,d_i} = \frac{1}{4}\sum_U (U\otimes U)\bar{\rho}^{\,i}_{\hat{A}_i\hat{B}_i|c_i,d_i}(U\otimes U)^\dagger, \tag{7}$$

where $\bar{\rho}^{\,i}_{\hat{A}_i\hat{B}_i}$ is defined in equation (6) and the sum is over $U \in \{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$.

It was shown in [BDSW96] that the resulting state is diagonal in the Bell basis $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$. In our notation, we get the following corollary:

Corollary 10. *For all $i \in [n]$ with $\mathrm{T_i} = 0$,*

$$\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_iC_iD_i} = \sum_{c_i,d_i} p(c_i, d_i)\,\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_i|c_i,d_i} \otimes |c_i\rangle\langle c_i| \otimes |d_i\rangle\langle d_i|, \tag{8}$$

where each $\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_i|c_i,d_i}$ is diagonal in the Bell basis. That is, it can be written as

$$\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_i|c_i,d_i} = \begin{pmatrix} \lambda_{\Phi^+} & & & \\ & \lambda_{\Phi^-} & & \\ & & \lambda_{\Psi^+} & \\ & & & \lambda_{\Psi^-} \end{pmatrix} \tag{9}$$

in the basis of the Bell states ordered as $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$ for some eigenvalues.

As in the projection step, the effect of applying a random unitary is restricted to the case $T_i = 0$ and can be done with only LOCC. The importance of this step lies in the following.

1. After the twirling step, the resulting two qubit state is a convex combination of Bell diagonal states (Corollary 10). This simplifies the analysis in the rest of the proof.
2. This step decouples $\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_iC_iD_i}$ from other registers while maintaining the entanglement of $\tilde{\rho}^{\,i}_{\hat{A}_i\hat{B}_i}$; see lemma 12 below for the formal statement. This property is crucial later on, in section 4.3.1.

4.1.3. Properties of the modified protocol

Protocol 2 can be described mathematically by an application of a sequence of maps one after the other. The maps describe both the actions of the verifier defined by the protocol as well as the actions of the measurement devices. We denote these maps by

$$\mathcal{N}_i\colon R_{i-1} \to R_i\hat{A}_i\hat{B}_iA_iB_iC_iD_iX_iY_iW_i. \tag{10}$$

The register R_{i-1} holds here the state of the uncharacterised devices in the beginning of the ith round. That is, R_{i-1} includes the state ϕ^i produced by the source as well as any other information kept by the measurement devices. R_i describes a register which can be used as internal memory for the devices.

The final state in the end of protocol 2 is denoted by

$$\tau = \tau_{\hat{A}\hat{B}ABCDXYTW}$$

and the state conditioned on not aborting by $\tau_{|\Omega}$.

We now present 3 statements regarding the modified protocol. All the proofs are rather simple and are given in appendix A.

For start, it follows from the definition of protocol 2 that the observed statistic of ρ (the state in the end of protocol 1) and τ (the state in the end of protocol 2) are the same. This, in particular, implies that the probabilities of aborting protocols 1 and 2 are identical.

Lemma 11. *The observed statistics and, hence, the probabilities of aborting protocols 1 and 2 are the same. That is,* $\rho_{ABXYTW} = \tau_{ABXYTW}$ *and* $\Pr[\neg\Omega]_\rho = \Pr[\neg\Omega]_\tau$.

Second, as mentioned in the previous section, the final state of protocol 2, τ, has the property that for any $i \in [n]$ the registers $\hat{A}_i\hat{B}_i$ are decoupled from the other registers. This can be shown using the structure of the state given in equation (8). The formal statement needed in the next sections is given in the next lemma.

Lemma 12. *Let* τ *denote the state after all rounds of protocol 2 (before conditioning on* Ω*). For every* $i \in [n]$ *let* K *include all registers different than* $\hat{A}_i\hat{B}_iC_iD_iT_i$ *such that we can write*

$$\tau_{\hat{A}\hat{B}ABCDXYTW} = \tau_{\hat{A}_i\hat{B}_iC_iD_iT_iK}.$$

Then,

$$H(\hat{A}_i|\hat{B}_iC_iD_iT_iK)_\tau = H(\hat{A}_i|\hat{B}_iC_iD_iT_i)_\tau.$$

The last thing which will be of use later is that $\rho_{|\Omega}$ is at least as entangled as $\tau_{|\Omega}$. That is, if one can distill $|\Phi^L\rangle$ from $\tau_{|\Omega}$ then one can also distill $|\Phi^L\rangle$, for the same L, from $\rho_{|\Omega}$. This follows from the fact that both modifications described above can be implemented using LOCC, and hence they cannot increase the entanglement. The formal statement is given in the last lemma of this section.

Lemma 13. *Let* Γ *denote the LOCC protocol used to distill* $|\Phi^L\rangle = \frac{1}{\sqrt{L}}\sum_{i=1}^{L}|i\rangle|i\rangle$ *from* $\tau_{|\Omega}$ *with error probability* ε. *Then, there exists another LOCC protocol* Δ *which can be used to distill* $|\Phi^L\rangle$ *from* $\rho_{|\Omega}$ *with the same error probability.*

Lemma 13 implies that instead of proving that there exists an entanglement distillation protocol starting from $\rho_{|\Omega}$ one can prove that there exists an entanglement distillation protocol starting from $\tau_{|\Omega}$. The advantage of doing this is that $\tau_{|\Omega}$ has some nice properties as was shown above. These will be of use in our proof given in the next sections. We emphasise again that protocol 2 acts just as a step in the proof. This is not a real protocol which we expect the verifier to implement in order to verify that the state produced by the source is highly entangled.

4.2. Single-round bound on the von Neumann entropy

The goal of the current section is to upper bound the conditional von Neumann entropy $H(\hat{A}_i|\hat{B}_i)$ of the states produced by the source and kept in the memory in protocol 2. As explained in section 2.3.2, a negative value of the conditional von Neumann entropy can be used to quantify the amount of entanglement present in the considered state. While there are many ways to quantify entanglement, the conditional von Neumann entropy is the one relevant for our proof technique (as will become clear below).

The first lemma of this section was already presented as lemma 5 in section 2.3.2. To simplify the equations in the lemmas below we use the notation of a Bell violation $\beta \in [2,\ 2\sqrt{2}]$ instead of the winning probability $\omega \in \left[\frac{3}{4},\ \frac{2+\sqrt{2}}{4}\right]$ (used in lemma 5). The two are related via $\omega = \frac{1}{2} + \frac{\beta}{8}$; the transformation back to ω will be done in the end.

Lemma 14. *For any Bell diagonal state* $\sigma_{\hat{A}_i\hat{B}_i}$ *as in equation (9) that can be used to violate the CHSH inequality with violation* $\beta \in [2,\ 2\sqrt{2}]$,

$$H(\hat{A}_i|\hat{B}_i)_\sigma \leqslant \mathscr{H}(\beta) - 1,$$

where

$$\mathscr{H}(\beta) = 2h\left(\frac{1}{2} - \frac{\beta}{4\sqrt{2}}\right) \tag{11}$$

and h *is the binary entropy function.*

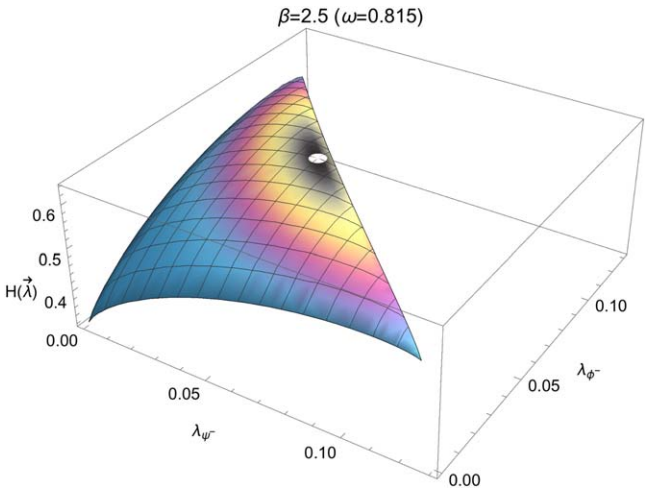

Figure 4. $H(\vec{\lambda})$ as a function of λ_{Ψ^-} and λ_{Φ^-}, for $\beta = 2.5$, in the region defined by the constraints of the optimisation problem stated in equation (31). As clearly seen in the plot, there is only a single maxima within the considered region. The white point denotes the solution given in equation (13).

The function $\mathscr{H}(\beta) - 1$ is plotted in figure 3 as a function of the CHSH violation β and the winning probability ω.

The proof is inspired by the work of [PAB+09], though we derive a bound on a different quantity. Neither result follows from the other. We give the proof sketch here; the full details are given in appendix B.

Proof sketch. First note that

$$H(\hat{A}_i|\hat{B}_i) = H(\hat{A}_i\hat{B}_i) - H(\hat{B}_i) = H(\hat{A}_i\hat{B}_i) - 1$$

since the marginal $\sigma_{\hat{B}_i}$ is a completely mixed qubit state (for a Bell diagonal state $\sigma_{\hat{A}_i\hat{B}_i}$).

We are left to upper bound $H(\hat{A}_i\hat{B}_i)$ which, given that the state is Bell diagonal, is simply the Shannon entropy $H(\vec{\lambda})$ of the probability distribution $\vec{\lambda} = (\lambda_{\Phi^+}, \lambda_{\Phi^-}, \lambda_{\Psi^+}, \lambda_{\Psi^-})$ defined by the four eigenvalues of σ[15]. Hence, our goal is to maximise $H(\vec{\lambda})$ under the constraint of having the correct Bell violation β.

Following [PAB+09], lemma 7 and [HHH95], equation (21) we translate the constraint on the violation of the state to the following constraint on the eigenvalues:

$$\begin{aligned}\beta = \max\{&2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^+})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2},\\ &2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^-})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^+})^2},\\ &2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Phi^-})^2 + (\lambda_{\Psi^+} - \lambda_{\Psi^-})^2}\}.\end{aligned}$$

Using the symmetry of our problem (e.g. under the exchange $\lambda_{\Psi^+} \leftrightarrow \lambda_{\Psi^-}$) one can simplify the above constraint and write the optimisation problem of interest as

$$\begin{aligned}&\max \quad H(\vec{\lambda})\\ &\text{s.t.} \quad \lambda_{\Phi^+} = \frac{1}{2}\left(1 - \lambda_{\Psi^-} - \lambda_{\Phi^-} + \sqrt{\frac{\beta^2}{8} - (\lambda_{\Psi^-} - \lambda_{\Phi^-})^2}\right)\\ &\quad \lambda_{\Phi^+}, \lambda_{\Psi^+}, \lambda_{\Phi^-}, \lambda_{\Psi^-} \geqslant 0\\ &\quad \lambda_{\Phi^+} + \lambda_{\Psi^+} + \lambda_{\Phi^-} + \lambda_{\Psi^-} = 1.\end{aligned} \tag{12}$$

The constraints of the optimisation problem imply that we can write $H(\vec{\lambda})$ as a function of only two variables, λ_{Ψ^-} and λ_{Φ^-}, for any value of β. As an example, $H(\vec{\lambda})$ is presented in figure 4 for $\beta = 2.5$.

One can solve this optimisation problem numerically; the solution is given by

$$\begin{aligned}&\lambda^*_{\Phi^+} = \left(\frac{1}{2} - \frac{\beta}{4\sqrt{2}}\right)^2; \quad \lambda^*_{\Psi^+} = \left(\frac{1}{2} + \frac{\beta}{4\sqrt{2}}\right)^2;\\ &\lambda^*_{\Phi^-} = \lambda^*_{\Psi^-} = \left(\frac{1}{2} - \frac{\beta}{4\sqrt{2}}\right)\left(\frac{1}{2} + \frac{\beta}{4\sqrt{2}}\right).\end{aligned} \tag{13}$$

[15] This observation was already made in [BBPS96]. In the context of [BBPS96], the eigenvalues $\vec{\lambda}$ are known. What we do here can be seen as an extension to the case where only the Bell violation is known and not the eigenvalues.

Now that we have a solution in hand, we can verify that it is indeed the correct local optimal solution. This can be done by taking the derivatives in the relevant directions and verifying that the point given in equation (13) is indeed a maxima. □

Next, we extend the claim of lemma 14 also to convex combinations of Bell diagonal states. This can be done easily using the definition of the conditional entropy. The proof is given in appendix B.

Lemma 15. *Let $\mathscr{H}(\beta)$ be as in equation (11). Then, for any state $\sigma_{\hat{A}_i\hat{B}_iC_iD_i}$ as in equation (8) that can be used to violate the CHSH inequality with violation $\beta \in [2, 2\sqrt{2}]$,*

$$H(\hat{A}_i|\hat{B}_iC_iD_i)_\sigma \leqslant \mathscr{H}(\beta) - 1.$$

Two remarks are in order.

1. The derived upper bound on $H(\hat{A}_i|\hat{B}_iC_iD_i)$ is tight; it is saturated by the Bell diagonal state defined via the eigenvalues given in equation (13).
2. As can be seen from figure 3, there is a regime of parameters ($\beta \lessapprox 2.2$) in which the conditional entropy is positive even though the state violates the CHSH inequality and, hence, is entangled. Indeed, it is known that some states, e.g. the Werner state, can be used to violate the CHSH inequality while presenting positive conditional entropy [FBB17]. This implies that the conditional entropy is not the optimal quantity to use when certifying entanglement in a DI manner. Yet, it is the relevant quantity when bounding the operationally distillable entanglement (using the known techniques) as we do below.

For the coming steps of our proof, we need an upper bound on

$$\sup_{\sigma\in\Sigma} H(\hat{A}_i|\hat{B}_iA_iB_iC_iD_iX_iY_i)_{\mathcal{N}_i(\sigma)},$$

where $\mathcal{N}_i$ describes the maps defining the rounds of protocol 2, as in equation (10), and Σ describes the set of possible states for which protocol 2 *does not abort*. On the conceptual level, the protocol does not abort when the observed frequency defined by the registers W imply that the average Bell violation is sufficiently high[16]. Thus, we shall now consider the set Σ defined as $\Sigma = \{\sigma|w(\sigma) \geqslant \omega_{\text{th}}\}$ where $w(\sigma)$ is the winning probability in the CHSH game of the state σ and ω_{th} is some threshold winning probability.

The following lemma can be proven using the above lemmas together with the definition of the conditional entropy and the transformation between β and ω; see appendix B for the proof.

Lemma 16. *For any $\omega_{\text{th}} \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$, let $\Sigma = \{\sigma|w(\sigma) \geqslant \omega_{\text{th}}\}$. Then,*

$$\sup_{\sigma\in\Sigma} H(\hat{A}_i|\hat{B}_iA_iB_iC_iD_iX_iY_i)_{\mathcal{N}_i(\sigma)} \leqslant (1-\gamma)\cdot g(\omega_{\text{th}}),$$

where

$$g(\omega_{\text{th}}) = 2h\left(\frac{1}{2} - \frac{2\omega_{\text{th}}-1}{\sqrt{2}}\right) - 1$$

and h is the binary entropy function.

4.3. Upper bound on the total smooth max-entropy

As mentioned in section 2.3.1, a lower bound on the one-shot distillable entanglement can be calculated using the conditional smooth max-entropy $H_{\max}^{\varepsilon_{\text{smo}}}$. The following quantitive relation was derived in [WTB17], proposition 21:

Lemma 17 ([WTB17])**.** *For any $\theta_{\hat{A}\hat{B}}$, $\varepsilon_{\text{dist}} \in [0, 1]$, and $\varepsilon_{\text{smo}} \in [0, \sqrt{\varepsilon_{\text{dist}}})$ there exists a one-way entanglement distillation protocol Γ, utilising classical communication from Alice to Bob, such that*

$$F(\Phi^L, \Gamma(\theta_{\hat{A}\hat{B}})) \geqslant 1 - \varepsilon_{\text{dist}},$$

where Φ^L is a maximally entangled state of Schmidt rank L for

$$\log L = -H_{\max}^{\varepsilon_{\text{smo}}}(\hat{A}|\hat{B})_\theta - 4\log\left(\frac{1}{\sqrt{\varepsilon_{\text{dist}}} - \varepsilon_{\text{smo}}}\right).$$

The objective of this section is, thus, to supply a negative upper bound on

[16] On the technical level we will need to consider some further details; see section 4.3.2.

$$H_{\max}^{\varepsilon_{\text{smo}}}(\hat{A}|\hat{B}ABCDXY)_{\tau_{|\Omega}}.$$

We do this using the EAT [DFR16]. The EAT gives a way of bounding conditional smooth min- and max-entropies in sequential processes in which certain systems of interest are being produced one after the other in overall n steps (not necessarily in an IID way). It, roughly, states that the total amount of smooth entropy accumulated during the *entire* process is n times the von Neumann entropy produced in a *single* step of the process. In our context, this translates to saying that the total amount of smooth max-entropy can be related to the von Neumann entropy considered in section 4.2. All the definitions and statements of the EAT which are necessary for our work are presented in section 3.4.

4.3.1. Prerequisites of the EAT

Before applying the EAT, we show that the prerequisites of the EAT hold.

Specifically, $\mathcal{N}_i$: $R_{i-1} \to R_i\hat{A}_i\hat{B}_iA_iB_iC_iD_iX_iY_iW_i$ are the channels defined by protocol 2 and will act as our 'EAT channels'. We need to show that these are indeed EAT channels, i.e. that they fulfil definition 6 (see section 3.4). To verify that this is the case note the following.

1. W_i are classical finite-dimensional systems. $d_{\hat{A}_i} = 2$ (due to the projection step of protocol 2).
2. The classical value W_i is a function of $A_iB_iX_iY_i$. Hence, it can be measured from the output of the EAT channels (for any input state) without modifying the state.
3. The necessary Markov-chain conditions hold, as stated in the next lemma.

Lemma 18. *For all $i \in [n]$ and any initial state,*

$$\hat{A}_1^{i-1} \leftrightarrow \hat{B}_1^{i-1}A_1^{i-1}B_1^{i-1}C_1^{i-1}D_1^{i-1}X_1^{i-1}Y_1^{i-1}E \leftrightarrow \hat{B}_iA_iB_iC_iD_iX_iY_i$$

holds for the final state τ of protocol 2.

The crucial ingredient in the proof is lemma 12, which asserts that the twirling step (Step 12 of protocol 2) decouples the states kept in each round from the other registers; see appendix C for the details. It now becomes clear why the twirling step is necessary—without it the required Markov-chain conditions do not hold. For example, one can imagine a source that creates two bipartite states ϕ^1 and ϕ^2 entangled *with one another*. In such a case the above Markov-chain conditions do not hold. Step 12 therefore *enforces* the necessary conditions while not destroying the entanglement between $\hat{A}_i$ and $\hat{B}_i$.

4.3.2. Max-tradeoff function

To apply the EAT we need to define a concave max-tradeoff function, as defined in definition 7. We construct one by following similar steps used to define min-tradeoff functions in [AFRV16].

Let p be a probability distribution over $\{0, 1, \perp\}$ resulting from the observed data w, i.e. $p(\tilde{w}) = |\{i|w_i = \tilde{w}\}|/n$ for $\tilde{w} \in \{0, 1, \perp\}$, and define

$$\Sigma^p = \{\sigma|\mathcal{N}_i(\sigma)_{W_i} = p\}.$$

Below we focus on probability distributions for which $p(0) + p(1) = \gamma$. The reason is that for $p(0) + p(1) \neq \gamma$ the set Σ^p is empty and the condition on the max-tradeoff function becomes trivial. For such p we can write $\omega = \frac{p(1)}{p(0)+p(1)} = \frac{p(1)}{\gamma}$.

Lemma 16 can now be used to define a max-tradeoff function for any p with $\frac{p(1)}{\gamma} \in \left[\frac{3}{4}, \frac{2+\sqrt{2}}{4}\right]$. Define a function f by

$$f(p) = \begin{cases} (1-\gamma)\cdot g\left(\frac{p(1)}{\gamma}\right) & \frac{p(1)}{\gamma} \in \left[0, \frac{2+\sqrt{2}}{4}\right] \\ \gamma - 1 & \frac{p(1)}{\gamma} \in \left[\frac{2+\sqrt{2}}{4}, 1\right], \end{cases}$$

where g is as in lemma 16:

$$g(\omega) = 2h\left(\frac{1}{2} - \frac{2\omega-1}{\sqrt{2}}\right) - 1$$

for h the binary entropy function.

From lemma 16 we have

$$\sup_{\sigma\in\Sigma^p} H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i)_{\mathcal{N}_i(\sigma)} \leqslant (1-\gamma)\cdot g\left(\frac{p(1)}{\gamma}\right),$$

and, hence, it follows that any choice of $f_{\max}(p)$ that is differentiable and satisfies $f_{\max}(p) \geqslant f(p)$ for all p will be a valid max-tradeoff function for our EAT channels.

For the final bound derived using the EAT to be meaningful we choose $f_{\max}$ for which $\|\nabla f_{\max}\|_\infty$ is finite (for f given above the derivative at $\frac{p(1)}{\gamma} = \frac{2+\sqrt{2}}{4}$ is infinite). Let

$$f_{\max}(p, p_t) = \begin{cases} f(p) & p(1) \leqslant p_t(1) \\ a(p_t)\cdot p(1) + b(p_t) & p(1) > p_t(1), \end{cases}$$

where p_t is a probability distribution over $\{0, 1, \perp\}$ and

$$a(p_t) = \frac{\mathrm{d}}{\mathrm{d}p(1)} f(p)|_{p_t} \qquad \text{and} \qquad b(p_t) = f(p_t) - a(p_t)\cdot p_t(1).$$

It follows from the definition of a and b given in the above equation that $f_{\max}$ is differentiable and, for any p_t, $\|\nabla f_{\max}(\cdot, p_t)\|_\infty \leqslant a(p_t)$. Furthermore, as f is a concave function, we also have that $f_{\max}(p) \geqslant f(p)$ for all p. Thus, $f_{\max}(p)$ is indeed a max-tradeoff function.

4.3.3. Applying the EAT

We are finally ready to apply the EAT, stated as theorem 8, to derive an upper-bound on the smooth max-entropy. The smooth max-entropy rate is governed by the following functions:

$$f(p) = \begin{cases} (1-\gamma)\cdot g\left(\frac{p(1)}{\gamma}\right) & \frac{p(1)}{\gamma} \in \left[0, \frac{2+\sqrt{2}}{4}\right] \\ \gamma - 1 & \frac{p(1)}{\gamma} \in \left[\frac{2+\sqrt{2}}{4}, 1\right], \end{cases}$$

$$f_{\max}(p, p_t) = \begin{cases} f(p) & p(1) \leqslant p_t(1) \\ \frac{\mathrm{d}}{\mathrm{d}p(1)} f(p)\Big|_{p_t}\cdot p(1) + \left(f(p_t) - \frac{\mathrm{d}}{\mathrm{d}p(1)} f(p)\Big|_{p_t}\cdot p_t(1)\right) & p(1) > p_t(1), \end{cases}$$

$$\eta(p, p_t, \varepsilon_{\mathrm{smo}}, \varepsilon_{\mathrm{snd}}) = f_{\max}(p, p_t) + \frac{1}{\sqrt{n}} 2\left(\log 5 + \left|\frac{\mathrm{d}}{\mathrm{d}p(1)} g(p)\Big|_{p_t}\right|\right)\sqrt{1 - 2\log(\varepsilon_{\mathrm{smo}}\cdot\varepsilon_{\mathrm{snd}})},$$

$$\eta_{\mathrm{opt}}(\varepsilon_{\mathrm{smo}}, \varepsilon_{\mathrm{snd}}) = \min_{p_t : \frac{3}{4} < \frac{p_t(1)}{\gamma} < \frac{2+\sqrt{2}}{4}} \eta(\omega_{\mathrm{exp}}\gamma - \delta_{\mathrm{est}}, p_t, \varepsilon_{\mathrm{smo}}, \varepsilon_{\mathrm{snd}}). \tag{14}$$

Lemma 19. *For any source and measurement device in the considered setting, let τ the state generated using protocol 2, Ω the event that protocol 2 does not abort, and $\tau_{|\Omega}$ the state conditioned on Ω. Then, for any $\varepsilon_{\mathrm{smo}}, \varepsilon_{\mathrm{snd}} \in (0, 1)$, either protocol 2 aborts with probability greater than $1 - \varepsilon_{\mathrm{snd}}$ or*

$$H_{\max}^{\varepsilon_{\mathrm{smo}}}(\hat{A}|\hat{B}ABCDXY)_{\tau_{|\Omega}} < n\cdot\eta_{\mathrm{opt}}(\varepsilon_{\mathrm{smo}}, \varepsilon_{\mathrm{snd}}),$$

where η_{opt} is defined in equation (14).

Proof. In sections 4.3.1 and 4.3.2 we showed that the prerequisites of the EAT hold and that $f_{\max}$ defined above is a max-tradeoff function for the EAT channels defined via protocol 2. The statement of the lemma then follows by applying the EAT, stated as theorem 8, with the above choices. □

4.4. Final statement

We are ready to prove our main theorem, theorem 4. For convenience, we restate the theorem here. $\mathcal{S}^{\mathrm{honest}}$ and $\mathcal{D}^{\mathrm{honest}}$ are as discussed in section 2.3.1 (or see section 4.4.1 below).

Theorem 4 (Main theorem). *For any $n \in \mathbb{N}_+$, $\varepsilon_{\mathrm{dist}}, \varepsilon_{\mathrm{snd}} \in [0, 1]$, and $\varepsilon_{\mathrm{smo}} \in [0, \sqrt{\varepsilon_{\mathrm{dist}}})$, protocol 1 is a DIEC protocol with the following.*

1. *Noise-tolerance (completeness): The probability that* P *aborts when applied on any* $\phi^{\text{honest}} \in \mathcal{S}^{\text{honest}}$ *using honest measurement devices from* $\mathcal{D}^{\text{honest}}$ *is at most* $\varepsilon_{\text{cmp}} \leqslant \exp(-2n\delta_{\text{est}}^2)$.

2. *Entanglement certification (soundness): For* any *source and measurement devices in the considered setting, either protocol 1 aborts with probability greater than* $1 - \varepsilon_{\text{snd}}$ *when applied on* ϕ *or* $E_D^{n,\varepsilon_{\text{dist}}}(\rho_{|\Omega}) \geqslant r$ *for* $r = \log(L)/n$ *and*

$$\log L = -n \cdot \eta_{\text{opt}}(\varepsilon_{\text{smo}}, \varepsilon_{\text{snd}}) - 4\log\left(\frac{1}{\sqrt{\varepsilon_{\text{dist}}} - \varepsilon_{\text{smo}}}\right),$$

where η_{opt} *is defined in equation (14).*

The theorem follows from the combination of lemmas 20 and 21 given below.

4.4.1. Noise-tolerance (completeness)

As the honest source and devices we choose to consider a source that produces identical and independent copies of a state $\phi^i = \sigma$ and measurement devices that apply the same measurements in each round when they are used. The state σ and the measurements are such that the winning probability achieved in the CHSH game is at least ω_{exp}. For example, one can choose $\mathcal{D}^{\text{honest}}$ to include the measurement devices that apply the optimal measurements performed in the CHSH game and the set $\mathcal{S}^{\text{honest}}$ to include all states $\phi^{\text{honest}} = \sigma^{\otimes n}$ for σ any noisy maximally entangled state that will result in winning probability $\geqslant \omega_{\text{exp}}$. The following lemma bounds the probability of protocol 1 aborting when using an honest device as above.

Lemma 20. *The probability that protocol 1 aborts for an honest implementation discussed above is at most* $\varepsilon_{\text{cmp}} \leqslant \exp(-2n\delta_{\text{est}}^2)$.

Proof. The protocol aborts in Step 11 when the sum of the W_i obtained during the test rounds is not sufficiently high (this happens when the estimated Bell violation is too low or when not enough test rounds were chosen). In the honest implementation the products $\chi(T_i = 1)W_i$ are IID RVs with $\mathbb{E}[\chi(T_i = 1)W_i] = \omega_{\text{exp}}\gamma$. Therefore, we can use Hoeffding's inequality:

$$\varepsilon_{\text{cmp}} = \Pr\left[\sum_i \chi(T_i = 1)W_i \leqslant (\omega_{\text{exp}}\gamma - \delta_{\text{est}}) \cdot n\right] \leqslant \exp(-2n\delta_{\text{est}}^2). \qquad \square$$

4.4.2. Entanglement certification (soundness)

Lemma 21. *For any source and measurement devices in the considered setting, denote by* $\rho_{|\Omega}$ *the state in the end of the DIEC protocol, protocol 1, conditioned on not aborting. Let* $\varepsilon_{\text{dist}}, \varepsilon_{\text{snd}} \in [0, 1]$ *and* $\varepsilon_{\text{smo}} \in [0, \sqrt{\varepsilon_{\text{dist}}})$. *Then, either protocol 1 aborts with probability greater than* $1 - \varepsilon_{\text{snd}}$ *or* $E_D^{n,\varepsilon_{\text{dist}}}(\rho_{|\Omega}) \geqslant r$ *for* $r = \log(L)/n$ *and*

$$\log L = -n \cdot \eta_{\text{opt}}(\varepsilon_{\text{smo}}, \varepsilon_{\text{snd}}) - 4\log\left(\frac{1}{\sqrt{\varepsilon_{\text{dist}}} - \varepsilon_{\text{smo}}}\right), \tag{15}$$

where η_{opt} *is defined in equation (14).*

Proof. Given a source and measurement devices in the considered setting, we first consider the hypothetical scenario in which *protocol 2* is being ran using the given devices. Putting lemmas 17 and 19 together we learn that either protocol 2 aborts with probability greater than $1 - \varepsilon_{\text{snd}}$ or there exists an entanglement distillation protocol Γ such that

$$F(\Gamma(\tau_{|\Omega}), |\Phi^L\rangle\langle\Phi^L|) \leqslant \varepsilon_{\text{dist}}, \tag{16}$$

for L as in equation (15).

According to lemma 13, equation (16) implies that there exists an entanglement distillation protocol Δ such that

$$F(\Delta(\rho_{|\Omega}), |\Phi^L\rangle\langle\Phi^L|) \leqslant \varepsilon_{\text{dist}},$$

with the same choice of parameters, i.e. L is as in equation (15). By the definition of the one-shot distillable entanglement, given in equation (2), we get that $E_D^{n,\varepsilon_{\text{dist}}}(\rho_{|\Omega}) \geqslant r$ for $r = \log(L)/n$.

Finally, lemma 11 tells us that if protocol 2 aborts with probability greater than $1 - \varepsilon_{\text{snd}}$ then protocol 1 aborts with probability greater than $1 - \varepsilon_{\text{snd}}$ as well.

Combining the above observations, the lemma follows. $\square$

The resulting distillable entanglement rates, $(\log L)/n$, are plotted in figure 2 as a function of the expected winning probability in the CHSH game ω_{exp} for different values of n. As seen from the figure, as the number of rounds of the protocol n increases our rate approach the optimal rate, for our proof technique, given by the IID asymptotic rate; see section 5 for further details.

We remark that one can also derive improved rates for finite number of rounds n by considering a slightly modified version of our DIEC protocol, similarly to what was done in [AFRV16], appendix B. As the modification of the protocol, the analysis, and the resulting rates follow directly by combining the analysis done here and that of [AFRV16], appendix B we do not present the details here.

5. Open questions

5.0.3. Tightness of our result

As mention in the previous sections, if one chooses to take the path of bounding the one-shot distillable entanglement using the smooth max-entropy, as done in the current work, then our quantitive results are tight to first order of n. However, this may not be the only way to go. Considering other proof techniques is crucial in order to achieve a result in which there is positive certified distillable entanglement regime whenever a violation of the CHSH inequality is being detected.

In general, it is known that there are Bell inequalities which can be violated by bound entangled states, i.e. entangled state which cannot be distilled [VB14] and, thus, for some Bell inequalities a zero rate regime, similar to the one observed here, is of a fundamental nature. However, for the CHSH inequality this is not the case [Mas06]: bound entangled states cannot violate the CHSH inequality. Hence, asymptotically, one should be able to certify distillable entanglement for any violation.

One way of assessing how far our results are from the optimal results, achievable using any proof technique, is to find *upper bounds* on the asymptotic distillable entanglement of the states achieving the highest conditional entropy given their Bell violation (see section 4.2). One possible starting point is to consider the sets of states described in [LDS17].

5.0.4. Possible extensions of our result

There are many possible ways of extending our work.

1. Our protocol and proof technique can also be modified to work with other Bell inequalities instead of the CHSH. This can potentially increase the rates when considering different types of honest sources of entanglement. For example, if one is interested in a source that emits partially entangled states then it probably makes more sense to consider the tilted CHSH inequalities [AMP12] rather than the CHSH. The only part of the proof which requiers a modification is the upper bound on the von Neumann entropy for a single round given in section 4.2. We remark that one can achieve such a bound for any Bell inequality for which a robust self-testing result is known, e.g. [BP15], combined with the continuity of the von Neumann entropy [Win16]. However, it is likely that such an approach will lead to relatively weak quantitive results. Thus, considering the von Neumann entropy directly for other Bell inequalities is a more promising direction.

2. An important direction to consider is the extension of our work to entanglement shared between more than two parties. This can then be used, for example, to consider scenarios and results as those derived in [Ban14, MPB+16] and extend them beyond the IID setting.

3. Another possible extension of the analysis done here is to consider DIEC protocols that employ the more general (but less fundamental) separability preserving operations [Rai97, CDKL01] rather than LOCC. To do so one should first consider one-shot distillation protocols which use separability preserving operations [BD11a].

4. One can also try to bound other operational measures of entanglement. For example, it will be interesting to lower bound the *one-shot entanglement cost*[17]. Such a bound can be viewed as a 'dual' of the one achieved in this work. Given the results of [BD11b], it is plausible that a similar proof technique as presented here can be used to achieve such a bound.

[17] In a related work [AFY17], a lower bound on the *entanglement of formation* of a quantum state (closely related to its entanglement cost) as a function of its Bell violation is derived. The considered setting and type of statement are different than the ones presented here and are not operational in our sense. For further details see [AFY17].

5. In a different direction, it can also be of interest to consider other settings than the one considered in the current work (as described in section 2.2). As different experiments may require different sets of assumptions, formulating other interesting scenarios and modifying the proof accordingly can be relevant.
6. Similarly, one may consider device-*dependent* and semi-DI versions of our work. For example, it is possible to study a one-sided DI scenario in which one of the measurement devices is completely characterised. The only part of our proof that needs be to modified in such a case is that given in section 4.2 while replacing Bell inequalities with Steering inequalities [CS16]. The rest of the proof will follow as is. The additional assumptions can potentially result in certification rates higher than the ones presented in the current work.

Acknowledgments

We thank Renato Renner, David Sutter, and Thomas Vidick for helpful discussions. We also thank Valerio Scarani for inviting RAF to visit his group at CQT Singapore, where the work on this project was initiated. RAF acknowledges support from the Swiss National Science Foundation (grant No. 200020-135048) via the National Centre of Competence in Research 'Quantum Science and Technology' and by the US Air Force Office of Scientific Research (grant No. FA9550-16-1-0245). JDB acknowledges support from the Swiss National Science Foundation (SNSF), through the Grant number PP00P2-179109.

Appendix A. Proofs of the lemmas in section 4.1

Lemma 9. *Consider a scenario in which the projection to the two qubit space is applied on the state ϕ^i directly after it is produced by the source in the* ith *round (i.e. before choosing the value of* $\mathrm{T_i}$*). Denote the resulting state in the end of the* ith *round in such a case by $\bar{\bar{\rho}}^i$. Then*

$$\bar{\bar{\rho}}^{i}_{\hat{A}_i\hat{B}_iA_iB_iC_iD_iX_iY_i} = \bar{\rho}^{i}_{\hat{A}_i\hat{B}_iA_iB_iC_iD_iX_iY_i},$$

where $\bar{\rho}^i$ is as defined in equation (6).

Proof. Firstly, for the rounds in which $T_i = 0$ there is clearly no difference between $\bar{\bar{\rho}}^{i}_{|T_i=0}$ and $\bar{\rho}^{i}_{|T_i=0}$. We show that the same holds also when $T_i = 1$. To this end we need to prove that

$$\bar{\bar{\rho}}^{i}_{A_iB_iX_iY_i|T_i=1} = \bar{\rho}^{i}_{A_iB_iX_iY_i|T_i=1}.$$

To see that this is indeed the case note that the successive application of the projections given in equation (5) and of the measurement as applied to create $\bar{\bar{\rho}}^i$ has the exact same effect as applying the measurement alone. Indeed,

$$\Pi^{A_i}_{a|x} = \sum_{c_i} \Pi^{A_i}_{a|x}\Pi^{A_i}_{c_i}$$

and similarly for Bob. Therefore, after tracing out the block registers C_i and D_i, both final states are identical.

As the probability for choosing $T_i = 0$ is, obviously, independent of when the projection is made the combination of the above statements implies the lemma. □

Lemma 11. *The observed statistics and, hence, the probabilities of aborting protocols 1 and 2 are the same. That is,* $\rho_{ABXYTW} = \tau_{ABXYTW}$ *and* $\Pr[\neg\Omega]_\rho = \Pr[\neg\Omega]_\tau$.

Proof. The only difference between protocols 1 and 2 is in the rounds in which $T_i = 0$. The observed statistics over $ABXYT$ and, hence, also W, depend however only on the rounds in which $T_i = 1$. Thus, $\rho_{ABXYTW} = \tau_{ABXYTW}$. The event Ω is defined according to the registers W and therefore the lemma follows. □

Lemma 12. *Let τ denote the state after all rounds of protocol 2 (before conditioning on Ω). For every $i \in [n]$ let* K *include all registers different than $\hat{A}_i\hat{B}_iC_iD_iT_i$ such that we can write*

$$\tau_{\hat{A}\hat{B}ABCDXYTW} = \tau_{\hat{A}_i\hat{B}_iC_iD_iT_iK}.$$

Then,

$$H(\hat{A}_i|\hat{B}_iC_iD_iT_iK) = H(\hat{A}_i|\hat{B}_iC_iD_iT_i).$$

Proof. We prove below that for all c_i and d_i we have

$$H(\hat{A}_i|\hat{B}_iT_iK,\ C_i = c_i,\ D_i = d_i) = H(\hat{A}_i|\hat{B}_iT_i,\ C_i = c_i,\ D_i = d_i). \tag{17}$$

Then, using the definition of the conditional entropy we write

$$\begin{aligned} H(\hat{A}_i|\hat{B}_i C_i D_i T_i K) &= \sum_{c_i,d_i} p(c_i,\, d_i) H(\hat{A}_i|\hat{B}_i T_i K,\, C_i = c_i,\, D_i = d_i) \\ &= \sum_{c_i,d_i} p(c_i,\, d_i) H(\hat{A}_i|\hat{B}_i T_i,\, C_i = c_i,\, D_i = d_i) \\ &= H(\hat{A}_i|\hat{B}_i C_i D_i T_i) \end{aligned}$$

and the lemma follows. We are therefore left to prove that equation (17) holds.

We split the proof to two parts corresponding to the different values of T_i. When $T_i = 1$ the register $\hat{A}_i$ has a deterministic value. Hence,

$$H(\hat{A}_i|\hat{B}_i K,\, C_i = c_i,\, D_i = d_i,\, T_i = 1) = H(\hat{A}_i|\hat{B}_i,\, C_i = c_i,\, D_i = d_i,\, T_i = 1) = 0. \tag{18}$$

The interesting case is thus $T_i = 0$. For $T_i = 0$ we can use the Bell diagonal structure of the state as given in equation (9). Consider the purification of the state $\tau_{\hat{A}_i\hat{B}_i|c_i,d_i}$:

$$\begin{aligned} |\tau\rangle^{c_i,d_i}_{\hat{A}_i\hat{B}_i F_i} = &\sqrt{\lambda_{\Phi^+}}\ |\Phi^+\rangle_{\hat{A}_i\hat{B}_i}|1\rangle_{F_i} + \sqrt{\lambda_{\Phi^-}}\ |\Phi^-\rangle_{\hat{A}_i\hat{B}_i}|2\rangle_{F_i} \\ &+ \sqrt{\lambda_{\Psi^+}}\ |\Psi^+\rangle_{\hat{A}_i\hat{B}_i}|3\rangle_{F_i} + \sqrt{\lambda_{\Psi^-}}\ |\Psi^-\rangle_{\hat{A}_i\hat{B}_i}|4\rangle_{F_i}. \end{aligned}$$

Note that K does not include the information encoded in F_i by the definition of K. Thus, including the register F_i we must have

$$\tau_{\hat{A}_i\hat{B}_i F_i K|C_i=c_i,D_i=d_i,T_i=0} = \tau_{\hat{A}_i\hat{B}_i F_i|T_i=0} \otimes \tau_{K|C_i=c_i,D_i=d_i,T_i=0}.$$

Moreover, we can freely trace F_i out while preserving the tensor product structure

$$\tau_{\hat{A}_i\hat{B}_i K|C_i=c_i,D_i=d_i,T_i=0} = \tau_{\hat{A}_i\hat{B}_i|T_i=0} \otimes \tau_{K|C_i=c_i,D_i=d_i,T_i=0},$$

from which it follows that

$$H(\hat{A}_i|\hat{B}_i K,\, C_i = c_i,\, D_i = d_i,\, T_i = 0) = H(\hat{A}_i|\hat{B}_i,\, C_i = c_i,\, D_i = d_i,\, T_i = 0). \tag{19}$$

The combination of equations (18) and (19), together with the definition of the conditional entropy, implies the lemma. □

Lemma 13. *Let Γ denote the LOCC protocol used to distill $|\Phi^L\rangle = \frac{1}{\sqrt{L}}\sum_{i=1}^{L}|i\rangle|i\rangle$ from $\tau_{|\Omega}$ with error probability ε. Then, there exists another LOCC protocol Δ which can be used to distill $|\Phi^L\rangle$ from $\rho_{|\Omega}$ with the same error probability.*

Proof. The only difference between protocols 1 and 2 is the addition of Steps 11 and 12.

These steps are such that their effect is restricted to the round in which they are applied. That is, whether we apply them or not in round i does not effect all other rounds $j \neq i$. In other words, they commute with the rest of the operations made in the protocol. We can therefore postpone them (for all steps i with $T_i = 0$) to the end of the protocol.

Furthermore, according to lemma 11 the additional steps do not change the observed statistics and the probability of the event Ω. Thus, we can also postpone them for after making the projection on Ω.

Denoting the combination of all the projections and rotations for all relevant rounds by the map Λ, the above means that the relation

$$\Lambda(\rho_{|\Omega}) = \tau_{|\Omega} \tag{20}$$

holds. Moreover, Λ can be implemented using only LOCC by definition of the Steps 11 and 12.

Let Γ denote the LOCC protocol used to distill $|\Phi^L\rangle$ from $\tau_{|\Omega}$ with error probability ε. We define

$$\Delta = \Gamma \cdot \Lambda \tag{21}$$

to be the successive application of Λ and Γ. As both are LOCC protocols, Δ by itself is an LOCC protocol. From equations (20) and (21) it follows that

$$\Delta(\rho_{|\Omega}) = \Gamma(\Lambda(\rho_{|\Omega})) = \Gamma(\tau_{|\Omega}).$$

Additionally, the operation Λ alway succeeds (i.e. equation (20) is always true). Therefore, Δ is an LOCC protocol which can be used to distill $|\Phi^L\rangle$ from $\rho_{|\Omega}$ with the same error probability as Γ. □

Appendix B. Proofs of the lemmas in section 4.2

Lemma 14. *For any Bell diagonal state $\sigma_{\hat{A}_i\hat{B}_i}$ as in equation (9) that can be used to violate the CHSH inequality with violation $\beta \in [2, 2\sqrt{2}]$,*

$$H(\hat{A}_i|\hat{B}_i)_\sigma \leqslant \mathscr{H}(\beta) - 1,$$

where

$$\mathscr{H}(\beta) = 2h\left(\frac{1}{2} - \frac{\beta}{4\sqrt{2}}\right) \tag{22}$$

and h is the binary entropy function.

Proof. First note that

$$H(\hat{A}_i|\hat{B}_i) = H(\hat{A}_i\hat{B}_i) - H(\hat{B}_i) = H(\hat{A}_i\hat{B}_i) - 1 \tag{23}$$

since the marginal $\sigma_{\hat{B}_i}$ is a completely mixed qubit state.

We are left to upper bound $H(\hat{A}_i\hat{B}_i)$ which, given that the state is Bell diagonal, is simply the Shannon entropy $H(\vec{\lambda})$ of the probability distribution $\vec{\lambda} = (\lambda_{\Phi^+}, \lambda_{\Phi^-}, \lambda_{\Psi^+}, \lambda_{\Psi^-})$ defined by the four eigenvalues of σ. Hence, our goal is to maximise $H(\vec{\lambda})$ under the constraint of having the correct Bell violation β.

Following [PAB+09], lemma 7 and [HHH95], equation (21) we translate the constraint on the violation of the state to the following constraint on the eigenvalues:

$$\begin{aligned}\beta = \max\{&2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^+})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2},\\ &2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^-})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^+})^2},\\ &2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Phi^-})^2 + (\lambda_{\Psi^+} - \lambda_{\Psi^-})^2}\}.\end{aligned} \tag{24}$$

To simplify the constraint we observe that both the objective function $H(\vec{\lambda})$ and the above constraint are invariant under the exchange of the eigenvalues with one another. Thus, we can assume without loss of generality that the optimal solution is restricted by, say, the first term in equation (24). That is, the eigenvalues of the state that maximise the entropy are such that

$$\begin{aligned}\beta &= 2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^+})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2}\\ \beta &\leqslant 2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^-})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^+})^2}\\ \beta &\leqslant 2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Phi^-})^2 + (\lambda_{\Psi^+} - \lambda_{\Psi^-})^2}.\end{aligned}$$

To see that this is indeed the case one can assume by contradiction that the second term in equation (24), and not the first one, is the one restricting the optimal solution. Then, by exchanging the values of λ_{Ψ^+} with λ_{Ψ^-} we get a different state, which is restricted by the first term instead of the second one, but attains the same value $H(\vec{\lambda})$. Hence, the solution defined by this exchange must be an optimal solution as well and we can work with it instead.

We can therefore restrict our attention to the optimisation problem

$$\begin{aligned}\max_{\vec{\lambda}} \quad & H(\vec{\lambda})\\ \text{s.t.} \quad & \beta = 2\sqrt{2}\sqrt{(\lambda_{\Phi^+} - \lambda_{\Psi^+})^2 + (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2}\\ & \lambda_{\Phi^+}, \lambda_{\Psi^+}, \lambda_{\Phi^-}, \lambda_{\Psi^-} \geqslant 0\\ & \lambda_{\Phi^+} + \lambda_{\Psi^+} + \lambda_{\Phi^-} + \lambda_{\Psi^-} = 1.\end{aligned} \tag{25}$$

The constraints stated in the optimisation problem given as equation (25) imply the constraint:

$$\lambda_{\Phi^+} = \frac{1}{2}\left(1 - \lambda_{\Phi^-} - \lambda_{\Psi^-} + \sqrt{\frac{\beta^2}{8} - (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2}\right). \tag{26}$$

Hence, $H(\vec{\lambda})$ can be written as a function of only two variables, λ_{Φ^-} and λ_{Ψ^-}, for any value of β.

Let us identify the region of interest in this plane. First, we have the linear conditions

$$\lambda_{\Phi^-}, \lambda_{\Psi^-} \geqslant 0 \tag{27}$$

$$\lambda_{\Phi^-} + \lambda_{\Psi^-} \leqslant 1. \tag{28}$$

Noticing that the optimisation (25) is invariant under exchange of $(\lambda_{\Phi^+}, \lambda_{\Psi^+})$ and $(\lambda_{\Phi^-}, \lambda_{\Psi^-})$, we can restrict our attention to solutions for which $(\lambda_{\Phi^-} - \lambda_{\Psi^-})^2 \leqslant (\lambda_{\Phi^+} - \lambda_{\Psi^+})^2$. Expressed in terms of our two independent variable λ_{Φ^-} and λ_{Ψ^-}, this condition is equivalent to

$$|\lambda_{\Phi^-} - \lambda_{\Psi^-}| \leqslant \frac{\beta}{4}. \tag{29}$$

This condition defines a diagonal band of interest in the space of λ_{Φ^-} and λ_{Ψ^-}.

Using the condition $\lambda_{\Phi^+} \geqslant 0$, we further restrict our region of interest:

$$\left(\lambda_{\Phi^-} - \frac{1}{2}\right)^2 + \left(\lambda_{\Psi^-} - \frac{1}{2}\right)^2 \geqslant \left(\frac{\beta}{4}\right)^2. \tag{30}$$

This condition describes the outside of a disk centred at $\lambda_{\Phi^-} = \lambda_{\Psi^-} = \frac{1}{2}$. For all $\beta > 2$, this identifies a unique region within the set defined by (27)–(29). Our objective function, the entropy $H(\vec{\lambda})$ is presented in figure 4 on our region of interest for $\beta = 2.5$.

The above optimisation can thus be re-expressed in terms of only two variables as

$$\begin{aligned}
&\max_{\lambda_{\Phi^-},\lambda_{\Psi^-}} \quad H(\vec{\lambda}) \\
&\text{s.t.} \quad \lambda_{\Phi^+} = \frac{1}{2}\left(1 - \lambda_{\Phi^-} - \lambda_{\Psi^-} + \sqrt{\frac{\beta^2}{8} - (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2}\right) \\
&\lambda_{\Psi^+} = \frac{1}{2}\left(1 - \lambda_{\Phi^-} - \lambda_{\Psi^-} - \sqrt{\frac{\beta^2}{8} - (\lambda_{\Phi^-} - \lambda_{\Psi^-})^2}\right) \\
&\lambda_{\Phi^-},\ \lambda_{\Psi^-} \geqslant 0 \\
&\left(\lambda_{\Phi^-} - \tfrac{1}{2}\right)^2 + \left(\lambda_{\Psi^-} - \tfrac{1}{2}\right)^2 \geqslant \left(\tfrac{\beta}{4}\right)^2.
\end{aligned} \tag{31}$$

One can solve this optimisation problem numerically within this region; the solution is given by

$$\begin{aligned}
\lambda^*_{\Phi^+} &= \left(\tfrac{1}{2} - \tfrac{\beta}{4\sqrt{2}}\right)^2; \\
\lambda^*_{\Psi^+} &= \left(\tfrac{1}{2} + \tfrac{\beta}{4\sqrt{2}}\right)^2; \\
\lambda^*_{\Phi^-} &= \lambda^*_{\Psi^-} = \left(\tfrac{1}{2} - \tfrac{\beta}{4\sqrt{2}}\right)\left(\tfrac{1}{2} + \tfrac{\beta}{4\sqrt{2}}\right).
\end{aligned} \tag{32}$$

The white point in figure 4 denotes this solution.

Now that we have a solution in hand, we can verify that it is indeed a correct solution (i.e. there are no numerical errors) which is locally optimal. To do so we need to check that the local gradient vanishes at our suggested solution in order to verify that it is indeed the maxima.

The objective function of (31) is invariant under the exchange $\lambda_{\Phi^-} \leftrightarrow \lambda_{\Psi^-}$. Its derivative in the direction $\lambda_{\Psi^-} - \lambda_{\Phi^-}$ must thus vanish at this point. To conclude about the local optimality of this solution, we are left to show that the derivative, on the curve $\lambda_{\Phi^-} = \lambda_{\Psi^-}$, vanishes at the point given in equation (32).

Under the constraint $\lambda_{\Phi^-} = \lambda_{\Psi^-}$, $H(\vec{\lambda})$ can be written as a function of the single variable, say, λ_{Ψ^-}. Explicitly, it reads

$$\begin{aligned}
H(\vec{\lambda}) = &-2\lambda_{\Psi^-}\log(\lambda_{\Psi^-}) - \left(\tfrac{1}{2} - \lambda_{\Psi^-} + \tfrac{\beta}{4\sqrt{2}}\right)\log\left(\tfrac{1}{2} - \lambda_{\Psi^-} + \tfrac{\beta}{4\sqrt{2}}\right) \\
&- \left(\tfrac{1}{2} - \lambda_{\Psi^-} - \tfrac{\beta}{4\sqrt{2}}\right)\log\left(\tfrac{1}{2} - \lambda_{\Psi^-} - \tfrac{\beta}{4\sqrt{2}}\right).
\end{aligned}$$

Thus,

$$\frac{\mathrm{d}H(\vec{\lambda})}{\mathrm{d}\lambda_{\Psi^-}} = \log\left(\left(\frac{1}{2} - \lambda_{\Psi^-}\right)^2 - \frac{\beta}{32}\right) - 2\log(\lambda_{\Psi^-})$$

and we have $\frac{\mathrm{d}H(\vec{\lambda})}{\mathrm{d}\lambda_{\Psi^-}}|_{\lambda_{\Psi^-}=\lambda^*_{\Psi^-}} = 0$ for $\lambda^*_{\Psi^-}$ as in equation (32), as we were set to show.

Therefore, we have identified a local optimum for our objective function. According to the plot given in figure 4, this local optimum is also a global optimum for equation (31) and hence equation (25). As the values of the optimal eigenvalues, given in equation (32), depend only on the CHSH violation β we can write $H(\hat{A}_i\hat{B}_i) = H(\vec{\lambda}^*) = \mathscr{H}(\beta)$ as in equation (22). Combined with equation (23) the lemma follows. □

Lemma 15. *For any $\beta \in [2, 2\sqrt{2}]$ let $\mathscr{H}(\beta)$ be as in equation (22). Then, for any state $\sigma_{\hat{A}_i\hat{B}_iC_iD_i}$ as in equation (8) that can be used to violate the CHSH inequality with violation β,*

$$H(\hat{A}_i|\hat{B}_iC_iD_i)_\sigma \leqslant \mathscr{H}(\beta) - 1.$$

Proof. We use the convex combination structure of $\sigma_{\hat{A}_i\hat{B}_iC_iD_i}$ given in equation (8). Let $\beta^{c_id_i}$ be the violation of the conditional state $\sigma_{|c_i,d_i}$. Then,

$$\begin{aligned} H(\hat{A}_i|\hat{B}_i C_i D_i) &= \sum_{c_i,d_i} p(c_i,\, d_i) H(\hat{A}_i|\hat{B}_i,\, C_i = c_i,\, D_i = d_i) \\ &\leqslant \sum_{c_i,d_i} p(c_i,\, d_i)(H(\beta^{c_i d_i}) - 1) \\ &\leqslant H\left(\sum_{c_i,d_i} p(c_i,\, d_i)\beta^{c_i d_i}\right) - 1 \\ &= \mathscr{H}(\beta) - 1, \end{aligned}$$

where the first equality follows from the definition of the conditional entropy, the second step holds due to lemma 14, in the third step we used the concavity of $\mathscr{H}(\beta)$, and the last step follows from $\beta = \sum_{c_i,d_i} p(c_i,\, d_i)\beta^{c_i d_i}$. □

Lemma 16. *For any* $\omega_{\text{th}} \in \left[\frac{3}{4},\, \frac{2+\sqrt{2}}{4}\right]$, *let* $\Sigma = \{\sigma | w(\sigma) \geqslant \omega_{\text{th}}\}$. *Then,*

$$\sup_{\sigma\in\Sigma} H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i)_{\mathcal{N}_i(\sigma)} \leqslant (1-\gamma)\cdot g(\omega_{\text{th}}),$$

where

$$g(\omega_{\text{th}}) = 2h\left(\frac{1}{2} - \frac{2\omega_{\text{th}} - 1}{\sqrt{2}}\right) - 1$$

and h is the binary entropy function.

Proof. We start with the following sequence of equalities that holds for any σ:

$$\begin{aligned} H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i)_{\mathcal{N}_i(\sigma)} &= \Pr[A_i,\, B_i = \perp]\cdot H(\hat{A}_i|\hat{B}_i X_i Y_i C_i D_i,\, A_i = \perp, B_i = \perp)_{\mathcal{N}_i(\sigma)} \\ &\quad + \Pr[A_i,\, B_i \neq \perp]\cdot H(\hat{A}_i|\hat{B}_i X_i Y_i C_i D_i,\, A_i \neq \perp, B_i \neq \perp)_{\mathcal{N}_i(\sigma)} \qquad (33) \\ &= \Pr[T_i = 0]\cdot H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i,\, T_i = 0)_{\mathcal{N}_i(\sigma)} \\ &\quad + \Pr[T_i = 1]\cdot H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i,\, T_i = 1)_{\mathcal{N}_i(\sigma)} \qquad (34) \\ &= (1-\gamma)\cdot H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i,\, T_i = 0)_{\mathcal{N}_i(\sigma)} \\ &\quad + \gamma\cdot H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i,\, T_i = 1)_{\mathcal{N}_i(\sigma)} \qquad (35) \\ &= (1-\gamma)\cdot H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i,\, T_i = 0)_{\mathcal{N}_i(\sigma)} \qquad (36) \\ &= (1-\gamma)\cdot H(\hat{A}_i|\hat{B}_i C_i D_i,\, T_i = 0)_{\mathcal{N}_i(\sigma)}. \qquad (37) \end{aligned}$$

To get equation (33) we used the definition of the conditional entropy. Equations (34) and (35) follow from the definition of protocol 2. Equation (36) holds since, by definition, the registers $\hat{A}_i\hat{B}_i$ have a deterministic initialisation value when a test is performed (nothing is kept in them) and hence $H(\hat{A}_i|\hat{B}_i A_i B_i C_i D_i X_i Y_i,\, T_i = 1)_{\mathcal{N}_i(\sigma)} = 0$. Finally, equation (37) follows from the fact that when $T_i = 0$ $A_i B_i X_i Y_i$ are all $\perp$.

The lemma then follows by combining equation (37) and lemma 15 while using the transformation $\beta = 8\omega_{\text{th}} - 4$ to replace β with ω_{th} in equation (11). □

Appendix C. Proofs of the lemmas in section 4.3

Lemma 18. *For all* $i \in [n]$ *and any initial state,*

$$\hat{A}_1^{i-1} \leftrightarrow \hat{B}_1^{i-1} A_1^{i-1} B_1^{i-1} C_1^{i-1} D_1^{i-1} X_1^{i-1} Y_1^{i-1} E \leftrightarrow \hat{B}_i A_i B_i C_i D_i X_i Y_i$$

holds for the final state τ *of protocol 2.*

Proof. The Markov-chain conditions can be written using the mutual information as

$$I(\hat{A}_1^{i-1}\colon \hat{B}_i A_i B_i C_i D_i X_i Y_i | \hat{B}_1^{i-1} A_1^{i-1} B_1^{i-1} A_1^{i-1} B_1^{i-1} X_1^{i-1} Y_1^{i-1} E) = 0.$$

Using the chain rule for the mutual information the above equation can be alternatively written as

$$\sum_{j\in[i-1]} I(\hat{A}_j\colon \hat{B}_i A_i B_i C_i D_i X_i Y_i | \hat{B}_1^{i-1} A_1^{i-1} B_1^{i-1} A_1^{i-1} B_1^{i-1} X_1^{i-1} Y_1^{i-1} E \hat{A}_1^{j-1}) = 0.$$

We now prove that for all $i \in [n]$ and $j \in [i-1]$

$$I(\hat{A}_j\colon \hat{B}_i A_i B_i C_i D_i X_i Y_i | \hat{B}_j A_j B_j C_j D_j X_j Y_j) = 0,$$

which, in turn, implies the lemma.

To this end, we use the following sequence of equations

$$\begin{aligned} H(\hat{A}_j|\hat{B}_j A_j B_j C_j D_j X_j Y_j) &\leqslant H(\hat{A}_j|\hat{B}_j C_j D_j T_j) \\ &= H(\hat{A}_j|\hat{B}_j C_j D_j T_j K) \\ &\leqslant H(\hat{A}_j|\hat{B}_j C_j D_j T_j A_j B_j X_j Y_j \hat{B}_i A_i B_i C_i D_i X_i Y_i), \end{aligned} \tag{38}$$

where K includes all registers of τ apart from $\hat{A}_j\hat{B}_j C_j D_j T_j$ such that we can write $\tau_{\hat{A}\hat{B}ABCDXYTW} = \tau_{\hat{A}_j\hat{B}_j C_j D_j T_j K}$. In particular, K includes $A_j B_j X_j Y_j \hat{B}_i A_i B_i C_i D_i X_i Y_i$. The first and last steps above are due to strong subadditivity of the entropy and the middle one follows from lemma 12.

Using the definition of the mutual information and equation (38) we get

$$\begin{aligned} I(\hat{A}_j\colon \hat{B}_i A_i B_i C_i D_i X_i Y_i|\hat{B}_j A_j B_j C_j D_j X_j Y_j) &= H(\hat{A}_j|\hat{B}_j A_j B_j C_j D_j X_j Y_j) \\ &\quad - H(\hat{A}_j|\hat{B}_j A_j B_j C_j D_j X_j Y_j \hat{B}_i A_i B_i C_i D_i X_i Y_i) \\ &\leqslant H(\hat{A}_j|\hat{B}_j T_j A_j B_j C_j D_j X_j Y_j \hat{B}_i A_i B_i C_i D_i X_i Y_i) \\ &\quad - H(\hat{A}_j|\hat{B}_j A_j B_j C_j D_j X_j Y_j \hat{B}_i A_i B_i C_i D_i X_i Y_i) \\ &= 0 \end{aligned}$$

and the lemma follows. □

References

[AFRV16] Arnon-Friedman R, Renner R and Vidick T 2016 Simple and tight device-independent security proofs arXiv:1607.01797

[AFY17] Arnon-Friedman R and Yuen H 2018 Noise-tolerant testing of high entanglement of formation *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)* (https://doi.org/10.4230/LIPIcs.ICALP.2018.11)

[AMP12] Acín A, Massar S and Pironio S 2012 Randomness versus nonlocality and entanglement *Phys. Rev. Lett.* **108** 100402

[Ban14] Bancal J-D 2014 Device-independent witnesses of genuine multipartite entanglement *On the Device-Independent Approach to Quantum Physics* (Berlin: Springer) pp 73–80

[BBP+96] Bennett C H, Brassard G, Popescu S, Schumacher B, Smolin J A and Wootters W K 1996 Purification of noisy entanglement and faithful teleportation via noisy channels *Phys. Rev. Lett.* **76** 722

[BBPS96] Bennett C H, Bernstein H J, Popescu S and Schumacher B 1996 Concentrating partial entanglement by local operations *Phys. Rev.* A **53** 2046

[BBS+13] Barreiro J T, Bancal J-D, Schindler P, Nigg D, Hennrich M, Monz T, Gisin N and Blatt R 2013 Demonstration of genuine multipartite entanglement with device-independent witnesses *Nat. Phys.* **9** 559–62

[BCP+14] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Bell nonlocality *Rev. Mod. Phys.* **86** 419

[BD10] Buscemi F and Datta N 2010 Distilling entanglement from arbitrary resources *J. Math. Phys.* **51** 102201

[BD11a] Brandao F G S L and Datta N 2011 One-shot rates for entanglement manipulation under non-entangling maps *IEEE Trans. Inf. Theory* **57** 1754–60

[BD11b] Buscemi F and Datta N 2011 Entanglement cost in practical scenarios *Phys. Rev. Lett.* **106** 130503

[BDSW96] Bennett C H, DiVincenzo D P, Smolin J A and Wootters W K 1996 Mixed-state entanglement and quantum error correction *Phys. Rev.* A **54** 3824–51

[Bel64] Bell J S 1964 On the Einstein-Podolsky-Rosen paradox *Physics* **1** 195–200

[BG11] Barrett J and Gisin N 2011 How much measurement independence is needed to demonstrate nonlocality? *Phys. Rev. Lett.* **106** 100406

[BNS+15] Bancal J-D, Navascués M, Scarani V, Vértesi T and Yang T H 2015 Physical characterization of quantum devices from nonlocal correlations *Phys. Rev.* A **91** 022115

[BP15] Bamps C and Pironio S 2015 Sum-of-squares decompositions for a family of Clauser-Horne-Shimony-Holt-like inequalities and their application to self-testing *Phys. Rev.* A **91** 052111

[BS16] Broadbent A and Schaffner C 2016 Quantum cryptography beyond quantum key distribution *Des. Codes Cryptogr.* **78** 351–82

[CDKL01] Cirac J I, Dür W, Kraus B and Lewenstein M 2001 Entangling operations and their implementation using a small amount of entanglement *Phys. Rev. Lett.* **86** 544

[CGS16] Coladangelo A, Goh K T and Scarani V 2017 All pure bipartite entangled states can be self-tested *Nat. Commun.* **8** 15485

[CHSH69] Clauser J F, Horne M A, Shimony A and Holt R A 1969 Proposed experiment to test local hidden-variable theories *Phys. Rev. Lett.* **23** 880–4

[CN16] Coudron M and Natarajan A 2016 The parallel-repeated magic square game is rigid arXiv:1609.06306

[Col09] Colbeck R 2009 Quantum and relativistic protocols for secure multi-party computation arXiv:0911.3814

[Col17] Coladangelo A 2017 Parallel self-testing of (tilted) epr pairs via copies of (tilted) chsh and the magic square game *Quantum Inf. Comput.* **17** 831–65

[CRSV16] Chao R, Reichardt B W, Sutherland C and Vidick T 2018 Test for a large amount of entanglement, using few measurements *Quantum* **2** 92

[CS16] Cavalcanti D and Skrzypczyk P 2016 Quantum steering: a review with focus on semidefinite programming *Rep. Prog. Phys.* **80** 024001

[CS17] Coladangelo A and Stark J 2017 Robust self-testing for linear constraint system games arXiv:1709.09267

[DFR16] Dupuis F, Fawzi O and Renner R 2016 Entropy accumulation arXiv:1607.01796

[DW02] De Wolf R 2002 Quantum communication and complexity *Theor. Comput. Sci.* **287** 337–53

[Ebe93] Eberhard P H 1993 Background level and counter efficiencies required for a loophole-free einstein-podolsky-rosen experiment *Phys. Rev.* A **47** R747–50

[FBB17] Friis N, Bulusu S and Bertlmann R A 2017 Geometry of two-qubit states with negative conditional entropy *J. Phys. A: Math. Theor.* **50** 125301

[GT09] Gühne O and Tóth G 2009 Entanglement detection *Phys. Rep.* **474** 1–75
[Hal11] Hall M J W 2011 Relaxed Bell inequalities and Kochen-Specker theorems *Phys. Rev.* A **84** 022102
[HHH95] Horodecki R, Horodecki P and Horodecki M 1995 Violating bell inequality by mixed spin-12 states: necessary and sufficient condition *Phys. Lett.* A **200** 340–4
[HHHH09] Horodecki R, Horodecki P, Horodecki M and Horodecki K 2009 Quantum entanglement *Rev. Mod. Phys.* **81** 865
[LDS17] Leditzky F, Datta N and Smith G 2017 Useful states and entanglement distillation *IEEE Trans. Inf. Theory* **64** 4689–708
[Mas06] Masanes L 2006 Asymptotic violation of bell inequalities and distillability *Phys. Rev. Lett.* **97** 050503
[MBL+13] Moroder T, Bancal J-D, Liang Y-C, Hofmann M and Gühne O 2013 Device-independent entanglement quantification and related applications *Phys. Rev. Lett.* **111** 030501
[McK16] McKague M 2016 Self-testing in parallel *New J. Phys.* **18** 045013
[MPB+16] McCutcheon W *et al* 2016 Experimental verification of multipartite entanglement in quantum networks *Nat. Commun.* **7** 13251
[NV17] Natarajan A and Vidick T 2017 A quantum linearity test for robustly verifying entanglement *Proc. 49th Annual ACM SIGACT Symp. on Theory of Computing* (New York: ACM) pp 1003–15
[AFRV98] Pfister C *et al* 2015 Understanding nature from experimental observations: a theory independent test for gravitational decoherence arXiv:1503.00577 [quant-ph]
[PAB+09] Pironio S, Acin A, Brunner N, Gisin N, Massar S and Scarani V 2009 Device-independent quantum key distribution secure against collective attacks *New J. Phys.* **11** 045021
[PR92] Popescu S and Rohrlich D 1992 Which states violate bell's inequality maximally? *Phys. Lett.* A **169** 411–4
[PV14] Plenio M B and Virmani S S 2014 An introduction to entanglement theory *Quantum Information and Coherence* (Berlin: Springer) pp 173–209
[Rai97] Rains E M 1997 Entanglement purification via separable superoperators arXiv:9707002
[Sca12] Scarani V 2012 The device-independent outlook on quantum physics *Acta Phys. Slovaca* **62** 347–409
[SW87] Summers S J and Werner R 1987 Bell's inequalities and quantum field theory: I. General setting *J. Math. Phys.* **28** 2440–7
[TCR10] Tomamichel M, Colbeck R and Renner R 2010 Duality between smooth min- and max-entropies *IEEE Trans. Inf. Theory* **56** 4674–81
[VB14] Vértesi T and Brunner N 2014 Disproving the peres conjecture by showing Bell nonlocality from bound entanglement *Nat. Commun.* **5** 5297
[VV12] Vazirani U and Vidick T 2012 Certifiable quantum dice *Phil. Trans. R. Soc.* A **370** 3432–48
[Wil10] Williams C P 2010 *Explorations in Quantum Computing* (Berlin: Springer)
[Win16] Winter A 2016 Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints *Commun. Math. Phys.* **347** 291–313
[WTB17] Wilde M M, Tomamichel M and Berta M 2017 Converse bounds for private communication over quantum channels *IEEE Trans. Inf. Theory* **63** 1792–817